DOCUMENT RESUME

ED 157 488 HE 010 338

TITLE	Evaluating Teaching in Higher Education. A Collection of Conference Papers.
INSTITUTION	London Univ. (England). Inst. of Education.
PUB DATE	Nov 75
NOTE	122p.
AVAILABLE FROM	The Secretary, UTMU, 55 Gordon Square, London WC1H ONT, England
EDRS PRICE	MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS	Case Studies (Education); *College Faculty; *College Instruction; Conference Reports; Course Evaluation; *Educational Accountability; Educational Assessment; Evaluation Methods; *Higher Education; Policy; *Teacher Improvement; Theories
IDENTIFIERS	*Great Britain

ABSTRACT

Some of the difficult problems related to introducing greater accountability for teaching were assessed and tentative policies considered at a British conference on higher education held in April 1975. R.C. Tress, Peter Kennedy, R. Rickett, Tom Driver, Marcus Merriman, and Alastair Stewart are the authors of the first set of papers entitled "National Policy on Evaluation." The second section, "The Theory and Practice of Evaluation" contains the following contributions: "The Evaluation of Teaching in Higher Education," by Roy Cox; "Evaluation of Teaching and the Institutional Context," by Carolyn Miller; "The Quality of Course Materials," by Michael Macdonald-Ross; "Is it Possible to Assess Teaching?" by Lewis Elton; "Student Evaluation of Teaching," by Colin Flood Page; "The Politics of Teacher Evaluation and Improvement: A Look at the United States," by Robert L. Woodbury; and "The Evaluation of Teaching--Some Comments," by Alex Main. The final section, "Case Studies," includes: "Student Perceptions of BSc Prelims. at Sussex" by Michael Eraut; "Report on a Study of Staff Perceptions of Promotion Procedures and the Role of the University Teacher," by Richard Startup; "Student Evaluation of Teaching at North East London Polytechnic," by Philip S. Bradbury and Paul Ramsden; and "Evaluation Studies of Six Courses in Geography, Electrical Engineering and Physics," by P.J. Black and R.J.D. Rutherford. (JMD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157488

Evaluating Teaching in Higher Education

A Collection of Conference Papers

HE 010 332

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UTMU

University Teaching Methods Unit

Produced by the University of
London Teaching Methods Unit/
printed by Page Bros (Norwich) Ltd/
cover design by Herb Gillman
© Institute of Education
ISBN 0 85473 053 2
First Published November 1975

This is one of a series
of publications produced by
the University of London
Teaching Methods Unit.
Details of these other publications
and/or matters concerning the
work of the Unit should be
addressed to The Secretary,
UTMU, 55 Gordon Square,
London WC1H 0NT

Contents

Preface ix

Introduction
Professor Sir Cyril Philips xi

Section I: National Policy on Evaluation

Dr R C Tress 1
Dr Peter Kennedy 8
Dr R Rickett 11
Tom Driver 15
Marcus Merriman 17
Alastair Stewart 19

Section II: The Theory and Practice of Evaluation

The Evaluation of Teaching in Higher Education
Roy Cox 22

Evaluation of Teaching and the Institutional Context
Carolyn Miller 34

The Quality of Course Materials
Michael Macdonald-Ross 39

Is it Possible to Assess Teaching?
Professor Lewis Elton 52

Student Evaluation of Teaching
Colin Flood Page 60

The Politics of Teacher Evaluation and Improvement: A Look at the United States
Robert L Woodbury 65

The Evaluation of Teaching – Some Comments
Alex Main 70

Section III: Case Studies

Student Perceptions of BSc Prelims. at Sussex
Dr Michael Eraut 73

Report on a Study of Staff Perceptions of Promotion Procedures and the Role of the University Teacher
Richard Startup 85

Student Evaluation of Teaching at North East London Polytechnic
Philip S Bradbury & Paul Ramsden 94

Evaluation Studies of Six Courses in Geography, Electrical Engineering and Physics
Professor P J Black & Dr R J D Rutherford 102

Contributors

Professor P J Black is a professor in the Department of Physics at the University of Birmingham and **Dr R J D Rutherford** is the Director of the Advisory Service on Teaching Methods at the same University

Phil Bradbury and **Paul Ramsden** are members of the Student Feedback Project at the North East London Polytechnic

Roy Cox is a senior lecturer in the University of London Teaching Methods Unit and was the organiser of the Conference on Evaluation of Teaching in Higher Education

Tom Driver is the General Secretary of the Association of Teachers in Technical Institutions

Professor L R B Elton is Head of the Institute for Educational Technology at the University of Surrey

Dr Michael Eraut is the Director of the Centre for Educational Technology at the University of Sussex

Dr Peter Kennedy is a senior lecturer in the Department of Physics at the University of Edinburgh and was a nominee of the Committee of Vice-Chancellors and Principals on the Co-ordinating Committee for the Training of University Teachers

Michael Macdonald-Ross is a senior lecturer in the Institute of Educational Technology at the Open University

Alex Main is Co-ordinating and Research Officer of the Co-ordinating Committee for the Training of University Teachers

Marcus Merriman is a lecturer in the Department of History at the University of Lancaster and was a representative of the Association of University Teachers on the Co-ordinating Committee

Carolyn Miller is a lecturer in education at the University of Sussex

Colin Flood Page is a lecturer in the Postgraduate School of Studies in Research and Education at the University of Bradford

Professor Sir Cyril Philips is the Vice-Chancellor of the University of London and Director of the School of African and Oriental Studies

Dr R Rickett is the Director of the Middlesex Polytechnic

Richard Startup is a lecturer in the Department of Sociology and Anthropology at the University of Swansea

Alastair Stewart is Deputy President of the National Union of Students and was a member of the Co-ordinating Committee for the Training of University Teachers

Dr R C Tress is the Master of Birkbeck College and is an Advisor to the Universities Authorities Panel

Robert L Woodbury is Associate Provost at the University of Massachusetts and at the time of the conference was a Visiting Scholar at the University of London Institute of Education

Preface

The papers contained in this volume were all read at a conference organised by the University of London Teaching Methods Unit and held in April 1975. The theme of the conference has been avoided until very recently in British higher education; now, however, with higher education expenditure coming under ever closer scrutiny, the relatively amateur status of the teaching role contrasts more sharply with the highly professional role of the researcher. Although staff development courses may partially redress the balance are they likely to produce important changes in the near future? The most promising way of ensuring that teaching may become more effective may be to encourage situations where staff learn more about what happens when they are teaching and are able to discuss their preparation and planning with sympathetic but critical colleagues, but is such an approach compatible with the formal requirements for probation suggested for universities and later supported in the Houghton Report.

Against this background the conference attempted to assess some of the many difficult problems related to introducing greater accountability for teaching, and in addition suggested tentative policies for consideration within the broad context of evaluation theory. The first section of the book is concerned primarily with the state of national policy on the evaluation of teaching in higher education. Various participants spoke of the efforts being made by particular institutions or groups, and despite the wide variety of institutions represented – from the Co-ordinating Committee for the Training of University Teachers and the Committee of Directors of Polytechnics through to the various unions concerned, AUT, ATTI, NUS – a considerable amount of agreement was reached.

The second section of the volume is concerned with the various academic papers that were presented to the conference. It is not necessary to comment in detail about them here, however, a few words of explanation must be provided about the paper written by Alex Main. Due to illness Mr Main was unable to attend the conference and read his paper, and subsequent pressure of work meant that only a shortened version could be produced. The final section of the book reports on four case studies in institutions where work on the evaluation of teaching is currently taking place, and the success of projects such as these is likely to be critical in judging the attempts to expand the evaluation of teaching in the future.

Finally, a brief note about the conference itself. Despite the limitations imposed on the event by a somewhat unsatisfactory venue, the majority of participants appeared very satisfied with the conference overall. A general evaluation of the conference was made, and the completed questionnaires revealed that a clear majority of those who attended thought that the two days had been useful and worthwhile.

Introduction
Professor Sir Cyril Philips

I am delighted to welcome you to this University and to this conference mainly because – to state the obvious – I am in no doubt that the subject is of the highest importance, and equally that during the next few years it will assume ever-increasing importance. Importance certainly, as I can see very clearly, for this federal University. I had the opportunity and pleasure yesterday of sitting in with our committee for non-teaching staffs which largely consists of secretaries of schools throughout the University and we were discussing there a policy paper on staff development and training. This referred to the development training not only of teachers in higher education but also of senior and less senior administrative officers, clerical officers and technicians. To be fully effective in a federal University we have to some considerable extent to see the whole of this training programme as a single programme, an integrated programme. It meets most obviously at the level of heads of departments, for it would be of little use generating a programme of study and assessment for heads of departments which is purely administrative or purely academic. Just as the two meet most obviously there, they also meet in many ways right through the hierarchy of those who are concerned in higher education.

I stressed that this is a federal University with special problems in training. For example, the Murray report on the University's future suggested that some of the administrative officers in Senate House should have a territorial or collegiate responsibility also. This is quite impracticable as a working system for we simply do not dispose of the resources to apply such a system. Equally, proposals for an interchange of staff between colleges, whether at the administrative or academic levels, is also very difficult to achieve in practice. More direct, more efficient, and, I would think, much more effective would be to look at the whole development of a federal University as seen in one sense through training. I have made my point that we sit in the centre of a University for which this subject is of the highest importance.

I have a personal concern also in this subject as the Head of a College and as a teacher. I have struggled very hard to interest what is a large academic staff in teaching and in the assessment of teaching – and with very uneven results. I have struggled with this over virtually a couple of decades, given a lot of time to it and my experience suggests that the Teaching Methods Unit and others interested in this activity really have an uphill task. My lowest point – and I mention this as an illustration – came when I went to spend a day with one of our departments. We had run over the whole question of what kinds of courses of training teachers might well enjoy, particularly what kind of in-service training should be introduced into the department and towards the end of the day one of the senior lecturers in the department, a person of very great experience and I would suspect a good teacher,

---

Note: This is the Introduction to the Conference given by Professor Sir Cyril Philips, Vice-Chancellor of the University of London, on 15th April 1975.

said, "I think, Director, you are wasting your time, none of all this is really necessary. There are no such things as principles of teaching." And I said, "Well, wait a moment, what about the simplest, you know, 'The Lecturer Should Be Audible'." "Not a bit, not a bit," the senior lecturer said, "not a bit. I worked under Malinowski and you could hardly hear Malinowski and as a consequence we all sat on the edge of our chairs to listen to him." I am not quite sure why this is so. It is of interest, for example, that later you will be looking at problems of accountability, the implication in part behind that being that external pressures on higher education institutions, the pressures of Parliament, the pressures of public opinion will increasingly cause teachers to take their professional role more seriously. I suppose this is true; and if such pressures are applied for example through national salary awards and such like measures I suppose they can have an effect. But this is to accept a sad state of affairs for we are by our contracts and by our profession dedicated to the role of teaching and should not therefore need any inducement whatsoever apart from our own personal professional pride and our own desire to reach the highest point of excellence as teachers. That should be the dominant note in any investigation of the kind which you are about to start into the role of teachers in institutions of higher education.

One of the more interesting experiences which I personally had in teaching was in my career in the war-time Army in which for a time I was in the Directorate of Training. And so I will close on a lighter note with an army anecdote. I have often been told in my own College that it would be very good if we had methods of assessment comparable with those applied in the Services, particularly, for example, that all reports should be shown to the individual on whom they are written. I call to mind a rather nice series of assessments in traditional Army hierarchical fashion which started with a report on a young officer by his Battalion Commander who wrote "I think that this young officer should go a long way"; and the report went up to the Brigade Commander who added, in a crisp style, "In my opinion he should go far." Then the Divisional Commander, not quite so sure, said, "I think, in fact, that he may be pushed on too far." And the Corps Commander, to whom this came, giving things a more practical twist said, "I am concerned that he won't go far enough." Finally the Army Commander wrote boldly across the bottom: "When can we start him on his way!"

I wish this conference well. I wish personally too, that I could attend, but I hope that I have indicated my own personal interest and, I trust, the interest of this University.

Section I: National Policy on Evaluation

Dr R C Tress

The fact that I have been identified and introduced in the context of salary negotiations of one kind or another sets the boundaries on the kind of contribution which I ought to make this morning. My heading is the *background* to issues of accountability for teaching, not the issues themselves, and I should stick to that, leaving others to take up the more interesting matters which will occupy you for the next two days.

My context is given by the fact that the papers which have been issued for the conference include the national agreement between the University Authorities Panel and the AUT on the procedure and criteria to be used in connection with probation and the probationary period (attached as Appendix 1). And the first thing for me to point out is that that agreement, as you will see, was first set moving in 1971 in what was then – although it is difficult to think of it in that context now – a major salary review. Part of the undertaking given to the government by the participants on the universities side was that, in view of certain of the improvements in academic salaries there would be a closer look at probation and efficiency bar reviews. Thus, the agreement is set in the salaries context and in the employment context. The universities have a pattern of employment now which can mean the employment of an individual for forty years, most of which is of guaranteed tenure, with an age of retirement which is later than for almost all public servants other than Her Majesty's Judges. What conclusion you draw from that I leave you to develop in the next two days. But there it is. Compared with civil servants who can be retired, and most of them do retire, at 60, the universities have a retirement age of 65 to 67.

So there are, in fact, two questions deriving from this employment situation of academic staffs, one of their capability and the other of their flexibility. I am going primarily to concentrate on the earlier of those two: capability, because my terms of reference are probation and probationary reviews. But it is worth reminding ourselves that in a way and, indeed, essentially all of us who are engaged in university teaching are under continuous assessment. We are under continuous assessment by our colleagues; we are under continuous assessment by our students. On the one hand, for many years during my teaching career I prided myself on the fact that, as a matter of principle, I always gave the first year's teaching. But towards the end of my teaching career I used to say, "Well maybe I had better hang on to this principle of first-year teaching before my colleagues demote me to the point when that is all that I am allowed to do anyway." One is being assessed all the time by one's colleagues and the older one gets the more the risks of senescence are apparent. On the other hand, one is also being assessed by one's students. Whenever I face the question of student assessment in talking to academic staff, I always ask them the question which of those who, having children at universities, have not at some time or other asked their own sons and daughters what they thought about the teaching they were getting and had taken what they said seriously.

But as I say, my prime context is that of probation and the probationary review. It took three years to work out the agreement concerning the procedure and you will observe that it touches only on probation and does not yet affect the matter of the efficiency bar. Personally, I don't think that matters a great deal. It is the probationary situation which is critical and the agreement records three elements in the probationary review procedure. There is first the review itself, shown in paragraph 3; there is secondly the safeguards against unreasonable termination, in paragraph 12; and there is thirdly a specific procedure for appeals, in the last paragraph of all.

Now, as far as the probationary review itself is concerned, there are two sets of conditions laid down: adequate performance and personal suitability. As regards adequate performance, there are in fact four components to it. There is adequate performance in teaching, in research, and in examining duties and administration *plus*, and this is the critical factor, sufficient promise, as spelled out in paragraph 6. And then there is a subsidiary paragraph 8 which takes note of the possibility – with many more safeguards written around it than around the other conditions – of uncooperativeness, of an irresponsible manner in laboratories and so on. A man can on that count not have his appointment maintained even although he shows both adequate performance in what he has been required to do and in intellectual terms a sufficient promise for the future.

The sole reference to assessors in the national agreement, I will point out, is in paragraph 7. It is that the probationer's head of department and other senior colleagues are the most experienced and skilled in assessing performance. The picking out of the sufficient promise component as well as the components of adequate performance, which are capable of more objective data, is one of the points which are emphasised in the agreement. So much for the probationary review.

As regards the safeguards against unreasonable termination, the agreement refers to the post-probationary confirmation of appointment as a decision of 'the employing university', that is to say it is a management function. It specifies that a probationer should be given an explanation for not confirming his appointment but it does not involve him or his representative in the process. Personal involvement is more specifically covered in the third component of the agreement: the appeal. There is specified in paragraph 13 the requirement that there should be a specific procedure for appeal to an appropriately constituted body deriving from the ordinances and statutes of individual universities where this is practical.

But for today's discussions, the probationary review itself is in the background. What you will be concerned with, I hope, is less the probationary review than the probationary period. In other words, the paragraphs in the national agreement to which I would most strongly draw your attention are paragraphs 4 and 11. These specify certain requirements during the period of probation. There are six points in this process. One, a co-ordinated development programme throughout the probationary period. Two, training of a helpful and comprehensive nature including (a) advice and guidance by a senior colleague, and (b) encouragement to attend formal courses of instruction. Three, the allocation of day-to-day duties and workload appropriate for a person of the probationer's age, standing and experience. Four, advice and warning of inadequacy given as early as possible and repeated as often as necessary. Five, a record of the probationary period including details of advice given full enough for it to be produced in evidence if required. And six, a written warning with a copy on the probationary record if the probationer's failings are sufficiently serious to justify consideration of non-confirmation of appointment. These are the six conditions which universities under the agreement are called upon to fulfil during the probation period.

Finally I would draw your attention to paragraph 11, on what happens if a university falls short of its responsibilities. There is a saving clause here. An employing university which declines to retain a person should be able to show that training in university teaching was made available, that continuing advice and help towards improvement were offered, and that due warning was given of inadequacies by the head of department or another responsible person. In other words, the university should be able to show that the six points which I covered have been met. A university, the paragraph 11 concludes, which is in serious default in respect of these retains the right not to confirm the appointment. But it may face arguments for the extension of the probationary period or even for confirmation of the appointment.

To sum up: the probationary agreement provides for a review procedure more specific, and indeed in some cases firmer, than has been operating in a good many universities hitherto. It lays on universities new responsibilities or responsibilities newly to be recognised – you may well say they were always there and should have been exercised previously – during the three-year probationary period. So whilst, as I said in the beginning, in our accountability for teaching we have all to take note that we are *always* under continuous assessment, up to the age of retirement, the responsibilities on universities regarding their new and younger members are now specifically stated. They should be adhered to and operated.

Appendix 1

Joint Report of the Universities Authorities Panel and the AUT Concerning the Procedure and Criteria to be Used in Connection with the Probationary Period

Introduction

The 1971 academic and related salaries settlement included the following agreement:

Lecturers
Probationary period to be 3 years with possible extension to 4 years in doubtful cases. Training procedures to be improved with thorough review prior to confirmation on the basis of revised and improved procedures and criteria.

In furtherance of this agreement a working party of Committee A has been engaged in reviewing the procedures and criteria for assessment of university teachers on completion of probation. The results of that review are set out below. *It should be stressed that the following paragraphs are intended to be applicable only to academic staff appointed to full-time established posts with tenure and to no others.* Furthermore anything said here is without prejudice to the rights of universities to terminate appointments for 'good cause' or for redundancy.

The Nature of the Probationary Review

1 The terms of probation as now laid down in salary agreements have certain precise implications which must be recognised:

(i) The probationary period is strictly limited, normally 3 years and at most 4 years. In practice and in order to allow a person not being retained time to look elsewhere for employment the experience upon which an assessment has to be based is normally not more than 2 years.

(ii) Within the probationary period a decision must be reached either that the appointment shall be terminated at the end of the probationary period or that the person concerned shall be confirmed in his appointment.

(iii) A judgement has therefore to be made about a person's promise on very limited evidence; and there are no 'in-between' positions. The probationary period may not be extended beyond 4 years.

(iv) Probation having been *satisfactorily* completed in one institution, it would not be appropriate for a second probationary period to be imposed by another institution to which the person concerned subsequently moved.

2 Thus, the decisions made on its probationary lecturers are among the most vital staffing decisions that a university can make. At this point, the ability and energy of a lecturer are both under scrutiny, with termination of employment the sanction against failure on either count. Thereafter, the university is committed by the confirmation of appointment to the judgement that its lecturers are capable at least to the standard for lecturers, which is the career grade. Within that grade, only their energy and application may thereafter be called in question. There are the incentives of promotion to a higher grade, of the regard of one's colleagues, of professional success and eminence. But the only sanction is a halt to salary increments at the efficiency bar except where a member of staff is found guilty of misconduct, incapacity or dereliction of duty under proceedings for termination of his appointment through 'good cause'.

3 In the circumstances, it is essential in the Working Party view that the decision to confirm an appointment at the conclusion of the probationary period must require a *positive act of decision* by the employing university. Such considerations would not apply to temporary appointments, that is, those for a stipulated period of years only.

Thus:

(i) while there should always be a definite prospect of confirmation of appointment whenever a probationary appointment is made (ie probationary appointments should only be made to established posts) there is no commitment on the part of the employing university to make confirmation of appointment at the conclusion of probation an automatic process;

(ii) at the end of the third probationary year the employing university should have offered confirmation of appointment or taken appropriate steps to end the appointment or offered an extension of the probationary period for a fourth and final year. Where confirmation or extension of probation is offered the probationer must have accepted the offer for it to be effective;

(iii) extension of the probationary period to a fourth year is without commitment on the part of the university to offer confirmation of appointment. At the end of the fourth probationary year if the appointment is not to be confirmed the appropriate steps must have been taken to end it.

(iv) A temporary appointment in place of an extension of a probationary appointment for a final year may be offered to allow the person concerned a longer period in which to find employment, he having been informed that a confirmation of appointment will not be offered.

Selection, Training and Development

4 The Working Party is of the opinion that universities must maintain high standards of selection procedures when they are considering making appointments to their academic staff. Where appointments have a probationary period it is incumbent on universities to provide training for the probationer of a helpful and comprehensive nature. Advice and guidance by a senior colleague nominated for this task, and encouragement to attend formal courses of instruction should be included. Attention should be paid to developments in the training of university lecturers at a national level as well as to internal courses of instruction. The probationer should receive a co-ordinated development programme which lasts throughout his probationary period and permits appropriate reports to be made, and remedial action to be taken where necessary, at regular stages. Universities should also ensure that the day-to-day duties and workload allocated to a probationer are appropriate for a person of his age, standing and experience.

Criteria

5 The primary consideration for the employing university in deciding whether or not to retain a person at the conclusion of his probation must be the long-term

interests of the university itself, of the other members of its staff, and of its students. But it is recognised at the same time that a university has a responsibility to assist the development of a member of staff in his probationary period; and also some concern for the future of a probationer whom it does not wish to retain. In the light of these general considerations, the Working Party has concluded that there are two sets of conditions which should be involved in any decision to offer confirmation of appointment. These are discussed in turn in the next four paragraphs.

6 The Working Party considers that the conditions first to be met should be as follows:

For a person to be offered confirmation of appointment following a period of probation, the employing university shall be satisfied that, having regard to his standing, age, experience and the opportunities he has been offered:

(a) he has satisfactorily engaged in the teaching of prescribed courses and the supervisory and tutorial work assigned to him;

(b) he has satisfactorily engaged in research towards the advancement of his subject;

(c) he has conscientiously carried out such examining duties and satisfactorily performed such administrative duties as have been required of him; and

(d) he shows promise by his work and enterprise of continuing to develop as a university teacher and scholar.

7 The first three of these conditions are fairly straightforward and each admit of objective data. The fourth condition is the more difficult in that, like the confirmation of appointment which is at issue, it is concerned with future prospects rather than with past performance. Assessment here depends inevitably upon subjective estimate – the estimate of the person's head of department and other senior colleagues. It is however in just this class of assessment, the assessment of academic worth and potential, that senior university teachers are most experienced and most skilled.

8 Secondly, some forms of behaviour, and they are very exceptional, may seriously harm the capability, good order or morale of a department, a faculty or even a university. For example, a person may be persistently unco-operative or careless or may deliberately make serious breaches in matters of confidentiality.

9 The Working Party accepts that such behaviour, if proved, may be a proper reason for not confirming a probationary appointment. However, judgements on these aspects of a person's overall academic performance can be more heavily influenced by subjective assessments and irrelevant considerations than judgements on the merit of the person's published work or teaching ability. In some cases, the faults

complained of in a probationary lecturer may be traceable, at least in part, to defects of management or personnel relations. Accordingly, it is essential that unfavourable reports on any matters of personal behaviour be made in such a way that the probationary lecturer has an adequate opportunity to make out his defence. It is further essential that wider consultations and broader enquiries take place than would be carried out in connection with a person's teaching or research ability.

10 The fact that a person holds or expresses views, on any matter, which differ from those of his colleagues or other members of the university is no reason for not confirming his appointment.

11 An employing university which declines to retain a person on grounds of inadequate performance or insufficient promise or personal unsuitability should be able to show (a) that training in university teaching was made available and (b) that continuing advice and help towards improvement were offered and due warning given of inadequacies by the head of department or other responsible person. As part of the 1971 salary settlement universities agreed to improve the standard of training provided for probationers. In the matter of continuing guidance it is important that heads of departments or other designated persons should be conscious of their responsibilities, and that advice and warning where warranted should be given as early as possible and repeated as often as necessary. It should be the responsibility of the university authorities to see that a record of the probationary period is kept. Such a probationary record must be full enough to enable it to be produced in evidence if required. In particular details of the advice given and of warnings given in respect of those matters described in paragraphs 5 and 8 above must be included in the record. Where the probationer's failings are sufficiently serious to justify consideration of non-confirmation of appointment a written warning must be handed to him and a copy placed in the probationary record. A university which is in serious default in respect of (a) or (b) above retains the right not to confirm the appointment on the grounds that the conditions set out in paragraphs 6 to 8 above have not been met, but it may face arguments for the extension of the probationary period or even for confirmation of the appointment.

Review Procedure

12 It is desirable to ensure that the offer of confirmation of appointment following probation is not unreasonably withheld. The Working Party considers that the review procedure should provide safeguards against unreasonable termination. It is the Working Party's view that an employing university should give an explanation of the reasons for not confirming probation to the individual concerned but that there should be no obligation to publish the details of the proceedings unless this is required by legal process.

13 The Working Party considers that there should be a specific procedure whereby

an individual may appeal to an appropriately constituted body against a decision not to confirm his appointment. This procedure should derive from the Ordinances and Statutes of individual universities where this is practicable.

Dr Peter Kennedy

I had agreed with Roy Cox to try and help at this conference before he told me that I was to do the 'institutional slot'. I am, I'm afraid, not a very institutional person and I doubt very much whether I shall be able to rise to the tasks which he has given to me in the programme. The institution in question is the Committee for the Training of University Teachers which is made up of four members of the Association of University Teachers, two members from the University Grants Committee, two persons from the National Union of Students and four representatives of the Committee of Vice-Chancellors and Principals of whom I am one. So you can see that this is really a very unlikely group to have an institutional view of the sort of things that we are talking about at this conference. We have taken it that our main role is to try to improve teaching and learning in the university sector, although we are now putting out feelers towards the polytechnics and so on. We have only nibbled at the corner of our task as yet and that is why, I suspect, we don't have firm views. We are really rather sold on the idea of diversity. Our hope is that all sorts of original things are happening in the various universities and, as a committee, we have little desire to advocate a national policy.

I will try to discuss the few institutional views we have later but I wonder if I might start off by talking about my personal views, although having seen the list of people here from the various institutes of education I am really rather worried about doing so. I am very much aware of my lack of expertise in this area. I personally got involved in this business by starting up our conference for new members of staff in Edinburgh and this colours my view about the whole problem very much.

I would like to draw a parallel between what we do on such courses and on the proposals which are being made at this conference. We started off with one general theme that we tried to run throughout our course, and it was that if we are to advocate something to new staff we must exemplify it in the course we give them. We do not ask them to sit in rows and then talk at them about small group teaching. We try not to discuss inaudibly the use of the voice in lectures. We tried to avoid audio-visual experts with illegible slides and so on. I am leading up to the fact that we talked a lot about evaluation on this course, mainly because we think evaluation is a dynamic and creative element in the teaching process. We offered the view that a teachers' values are expressed by the sort of tests that he sets and that this test can be used to shape the learning process. We believe that it is unfair and unjust to undergraduates to advocate something like 'wide reading' or 'critical thought' and then to test in our examinations some trivial level of recall. We try to encourage our new staff to make their testing procedures open and clear.

Having discussed the 'why' of assessment, we talk a little about the 'how' to do it. We talk about how to make tests reliable and about their validity and so on. We have been saying these things to new staff about undergraduate evaluation and I want to suggest that, as employers, we must practice what we preach. If we are to evaluate the teaching of our staff our methods must be creative and not inhibiting or hurtful. Let me consider first the 'why' of the problem. Why do we want to evaluate teaching or teachers? I do hope that we are not thinking of ranking our teachers, that we are not giving first- and second-class certificates and so on. We may, however, have to think in terms of *failure*. It occurs to me that, if we don't select our staff in any way for their teaching ability and at the end of their probationary period the probability of failing for teaching reasons is vanishingly small, then I am not sure that anyone will think that we are taking this teaching business very seriously.

I would like to return for a moment to the Agreement on probation. As I read it, it stresses the fact that the AUT and the universities have a common purpose, a common interest in using the probationary period as an apprenticeship in which some skills are acquired and the 'right' attitudes to a teaching career can be encouraged. We are really interested in 'teaching effectiveness', and this aspect leads me away from too much emphasis on the individual because a person's teaching effectiveness depends so much on what he is permitted to do and on the environment in which he is asked to teach. I hope that this Agreement may well bring us a little more out into the open so that we see the whole teaching process as a combination of teachers and students in a common enterprise. I hope the assessment of the individuals won't be overdone. So 'why' should we assess teaching? To the extent that such assessment in the universities is going to tell us how well or how badly we are doing, I think we should do it, and we should encourage our new staff to recognise that that is 'why' we are doing it.

The 'how' of doing it is going to be of great importance. I hope we won't, in fact, test trivial things. It was suggested to me once that one way of telling if a chap was good at teaching, was to count the number of slides he had made. Now that seems to me a bizarre suggestion. The development of non-trivial creative methods of evaluating teaching innovations seems to me critical. Certainly our assessment method must be open and clear to the people being assessed. I think that this is going to throw a very difficult and important role on to the department. They must make explicit to new staff what it is that the department requires of them, where they are failing if they are and how they can improve. I don't know who is going to do this, it may be kindly senior members, but someone in the department is going to have a very difficult task.

Can we make this element of assessment of our teaching creative in the way we advocate it for our undergraduates? I am hoping very much that from this conference I shall learn how. One last point which is a personal one for me but I think it may be more general. I believe that the people involved in running conferences for our new staff should not be the assessors. There would be a loss of trust if we who run conferences were the same men who later became responsible for making these assessments. I think these roles must be separated. We then must ask who is going to do it and that is also a very difficult problem.

Those are my own views, and I would like now, if I may, to try to suggest how many of these might be held by the Committee. There are others present from that Committee and they can disagree with me later. I think our institutional views will seem very low level because of the decision we have taken not to thrust things on to universities. At the moment, we don't believe that there are really reliable methods of testing teaching effectiveness. At the end of these two days we may change that belief. We are convinced that a professional approach and careful planning in teaching pays off. We feel strongly that if we encourage good teaching and good learning in the universities we also encourage an element of *self-assessment* and that this is the really important thing to nurture. To the extent that there should be institutional assessment there must be a trusted method of evaluation and we all know that is the rub. Are our departmental teachers going to trust the assessment methods we are suggesting? One danger, you will all be aware of, is of the damage that can be done by an activist. Someone who comes into a department or university and tries to produce sudden change, or produces expectation of sudden change. He is, I think, largely counter-productive.

The sort of thing that our Committee could do, if you asked us, is to find out what happens at the moment. Dr Tress was suggesting that in fact assessment of teachers goes on all the time and this is true. We could, perhaps, find out to what extent it is formal, try to find out the best practice and, perhaps, propagate this through our university contacts. We have an important function to support those poor chaps in the departments who are being asked to assess their colleagues and write things down about them and say things to them. They are going to be in a highly exposed position and it is up to our committee to send out support to help them in this difficult task. We could perhaps bring over an elder statesman from America with some of the American experience, if this were thought to be valuable. We are tremendously aware of the dangers and the importance of timing in these things.

May I return to a personal note for a moment? I think there has been good teaching in universities for years. There always has been, but to some extent when I first started it was regarded as a slightly endearing aberration if someone showed strong interest in teaching. I think we have seen in the last few years a change from this situation to one in which it is seen as our professional duty to try and make our teaching effective. This may be due to the courses that we have run and it may not, but they have had some effect. What I am convinced of is that the imposition of some kind of assessment, or even the over enthusiastic advocacy of an injudiciously chosen one, can really do a great deal of harm to the slow process we have observed. If we are going to evaluate our teachers and our teaching we must be very careful, and this is where – for me at least – this conference has so much to offer because I am hoping it may help to prevent us from making this kind of mistake.

Dr R Rickett

My presence here today may derive from the conference organiser's view that wherever higher education topics are discussed then there must be the obligatory polytechnic representative. In other spheres we know sometimes the obligatory woman, the obligatory overseas student, the obligatory trade union representative. I am here, I suppose, to some extent as the obligatory poly-man. It is usually felt that if one can say something that puts the polytechnics in a very different light from the universities, so much the better. On another occasion I would be delighted to explore the rationale of the binary system, but I do not believe this is the occasion to do so. Let it suffice to say that I have come here not to give any support to the idea that polytechnics are concerned with the development of high standard teaching whereas universities are not, any more than I would accept the proposition that polytechnics are less concerned with the quality of research than are the universities. In fact, the idea that good teaching must be at the expense of research or vice versa is not in accord with my experience.

Since I am the obligatory polytechnic director of this conference I think it might be of interest and of relevance for you to know how I was your man for the day. I was, for want of a better word, chosen to address you because one of the sub-committees of the Committee of Polytechnic Directors received an invitation outlining the purposes of the conference. On this occasion there was a noted reluctance on the part of my colleagues to put their names forward and they are not a batch of people normally noticed for their modesty. I therefore made a little speech drawing attention to the fact that here was a conference discussing the most important commodity of any polytechnic or other institution of higher education: that of the teaching force, and talking about its quality and the control of that quality, and yet we are somewhat reluctant to agree to participate. I found that this little speech had been treated as an election speech and I was unanimously elected.

I have no special qualification in these matters of teacher assessment. Therefore it will be no part of my talk today to attempt to delve into areas where I have no professional competence. In any case, in reading the supporting papers of the conference especially that of Mr Cox, it is apparent that all my very original ideas had already been anticipated and I would take the opportunity to say how well those papers were written. My brief contribution is therefore concerned with the attitude and role of the directors towards this whole business of teaching assessment. When I say directors, I must make it quite clear, that I don't regard polytechnics as a generic term, and so if for no other reason I must state that the views I put forward today in no way commit any other polytechnic director. The reason for telling you how I was elected to be here today is relevant to the problems we are discussing.

Although as I said earlier, directors are normally quite happy to discuss any topic on higher education, it is interesting that they were somewhat reluctant to be involved in this one. Not, I think, because they think it is less than important, but they are worried about advancing views in an area of such great importance: the educational process, which is so lacking in quantitative methods or clear unequivocal guidelines. In fact, the contributory element in giving such prominence to research achievement in promotion and career development in higher education possibly derives from the fact that it is dependent on published work which is relatively easy to assess. Indeed some people have advanced the view that it can even be quantitatively assessed by weighing or counting. Many speakers today and tomorrow will, no doubt, draw attention to the lack of an agreed method of assessment and it will also be suggested that when part of the method is agreed, what you might call the assessment proper, the conclusions drawn from the assessment may be unsatisfactory or equivocal. I would not want to minimise those dangers. However, the conference should not shy away from its responsibility in giving guidance, especially when the supporting papers would suggest that worthwhile objective assessment of teachers in higher education can be made. My personal contribution to this conference is to ask you to recognise that professionally sound methods of assessing teachers' performance will require a system to be adopted within institutions that allows account to be taken of them. In other words it will be not enough for universities and polytechnics to be in favour of such developments unless they adopt forms of organisation which allow this assessment to be made. What are the characteristics of such an organisation? First, they must allow for maximum objectivity. Secondly, they must be systematic over a sufficiently long period and implemented formally. Thirdly, they must be reflected in the staff development pattern adopted and the career prospects of the staff.

Whatever the case for assessment has been in the past, it is now quite clear with the size of universities and polytechnics and the prospects for developing even larger institutions, that it is very important now. I would add that it is even more important when we are dealing with inter-disciplinary courses where staff are teaching over a range of courses and we need information on how they teach wherever they teach. Other considerations include major developments in learning methods and in the case of polytechnics, the impact of Houghton and the subsequent Burnham agreement, which requires polytechnics to actively consider research when making appointments and promotions. Now, if we are going to actively consider research and allow for an equal consideration to be given to teaching, we must adopt, as I said, objective and systematic methods of understanding what the teaching role is and how well it is being carried out. I would say in echoing a previous speaker that in this

Figure 1 Diagrammatic Representation of part of the Academic Structure of Middlesex Polytechnic

Director

Deputy Director

Assistant Directors – Physical Resources – Learning Resources – Student Affairs – Design (Research and Consultancy)

Course leaders Resource Centres	Dr E Booth (VI) Eng Degree courses, etc.	T Dockerty (VI) BSc Civil Eng	R Kitchener (VI) BSc Soc & Tech	Dr M Edwards (VI) Social Work courses, etc.	Dr J Farquhar (VI) BA Social Science	A Hindmarch (VI) Business Studies Degree courses, etc.	N Sida (VI) Business and Management short courses	R Waterhouse (VI) BA Humanities	G Younson (VI) DipAD Fine Art	R Marber (VI) DipAD Graphic Design	N Thomas (VI) DipAD Textiles/Fashion	N French (VI) DipAD 3 D Design	All other course leaders V, IV, III, PL or SL/III
Dean of Engineering, Science and Mathematics J Osborne-Moss Grade VI+ D Humphris (IV+) Dr J Golden (PL) A Findlay (PL)													
Dean of Social Science A Holt Grade VI+ E Lewis (PL) R Aldridge-Morris (PL) N Lee (PL)													
Dean of Business Studies and Management To be appointed Grade VI+ A Syme (VI) P Helps (PL) Dr R Howells (Reader)													
Dean of Humanities and Education To be appointed Grade VI+ R Gunter (V) P Leary (PL) T Robb (VI) B Davies (PL)													
Dean of Art and Design To be appointed Grade VI+ L Roberts (IV) P Green (IV+)													

connection it is essential to avoid over simplistic views of assessment. I would go so far as to say, I think, that it would be better that we did not attempt the exercise at all, than we did it in a way which undermined the confidence of the people in it and so detract it from the total educational process.

At the Middlesex Polytechnic an attempt has been made to cater for a serious approach to the assessment of teaching. We have adopted a matrix form of organisation to achieve this. It was also an attempt to deal with a lot of other problems and an outline of the system is given in Figure 1, from which you will see that the problems dealt with by the various elements in the matrix do not only deal with teacher assessment. Along the top are the various courses which relate possibly more closely to departments in other universities or polytechnics in which there is a head of department, here it is the head of a course, or course leader who is responsible for the degree whilst the staff are assigned to the various resource centres. It is the dean's responsibility to assess the staff and to do this in an extremely systematic way. Every staff member is interviewed at least once a year, by the dean or any other resource centre personnel, usually three or four as you see, and preferably with another member of the resource centre present or even a member of another resource centre. Before the staff member attends for interview he is given a form which gives the full details of his career to date including courses he is teaching or has taught, short courses he has been on, research he or she is doing and any other relevant information including qualifications obtained, so that they come to the interview fully briefed on what is on their file about them. They have full access to it, and the interview takes place and is called a staff development interview with all the aspects of the man's future and present career being discussed and the opportunity given for the man himself to make a major contribution to the interview. Before the interview the dean or resource centre personnel will have had the chance to consult with the course leaders of the courses on which the member of staff is teaching so that a fully worked out picture and a pattern of development is possible. As a result of this interview opportunity may be given to diversify into different lines of research, or to take other courses, or if there is any deficiency in teaching to go off and remedy it in some specific way. For new staff who are on probation it is the responsibility of the dean to organise the supervision of their teaching and for them to have a reduced teaching programme. Any suggestion of new course material or new course developments which a staff member wishes to be involved in can be dealt with on the occasion of the interview. And therefore a quite specific picture is being built up of the man, together with a formalised way of giving advice. Thus when the problem arises concerning his future career, this can be dealt with in a thoroughly formal way and any weaknesses identified and corrected in order that opportunities for improvement or advancement can be taken. I stress that this is only one way of approaching staff development and I have had to give you very much a thumb-nail sketch.

There is a danger that in discussing the whole business of teaching and teaching assessment that the pendulum will swing away from the need for a scholarly approach to the subject, and in anything that I have said I hope that this is not implied because I am certain that a key aspect of any teacher is that he is thoroughly conversant with the subject which he chooses to teach. Without this whatever advances are made in teaching method the student will certainly lack confidence in what is being taught and how it is being taught. In fact, Mr Cox in his paper refers somewhere to the possibility of teachers being regarded as learning managers. I do not shy away from that prospect, in fact the Open University has taught us that in many respects this can be made a very desirable approach. However, this is not to suggest that the quality and level of teacher and certainly the scholarship with which he approaches his subject need be any less. But the interlinking of the teaching and the learning process is absolutely vital and therefore I think this is an extremely valuable conference, and I will be interested to see the outcome. I am certainly going to use the papers that have been presented so far for distribution within my own institution because already I am sure that they are extremely valuable.

Tom Driver

I am a Trade Union official and it is my job to deal with the problems that arise as a result of the type of discussions that you are having today and the consequences and reactions you draw from them. I taught for over thirty years and am familiar with the other end as well. It seems to me that far too much attention is being given to problems of assessment and not nearly enough to the problem of assistance. There are two points at which assessment is a particular problem for teachers in higher education and in which one becomes involved. One is the problem at the end of a period of probation, if there is one, and secondly at the efficiency bar, if there is one. These are points which affect tenure, salary and so on and are therefore particularly sensitive points. But it seems to me that the question of the quality of teaching is equally important whether you have gone through a probationary period or an efficiency bar or whatever, and it is a continuing problem to which attention has to be given throughout the life of a teacher. Too much lip service is paid to producing good quality teachers. Teachers' unions have two functions to perform. One is to protect their members. Very occasionally the best interest of their member is that he ceases to be a teacher, although this occurs very rarely. The number of cases that one has to deal with, certainly that I have had to deal with in a fairly long experience of teaching unions, is so small as to suggest that talk of an assessment system is getting things out of proportion. You deal with the cases as they arise and they are not very great in number. The second thing a teachers' union has to do is to bring about the conditions in which good quality teaching can take place and in which that quality can be improved.

If you take the further education side with which I am particularly concerned, everybody agrees that teacher training is a good thing. The fact, of course, is that very little was done about it for a long time. That was not due to lack of goodwill by the teaching staff or even by the local authorities, but has been a governmental dictat. For in 1966 there was a report which recommended compulsory professional training for all new teachers in certain grades which, I think, achieved a record. Within an hour of publication it had been turned down by the government. The government normally take several months or several years before they turn a report down, or before they approve it, and then they approve it in such a way that nothing happens. At long last, however, it looks as though, thanks to the activities of the Advisory Council for the Supply and Training of Teachers, professional training for those who are not already professionally trained before coming into further education will be a contractual requirement within the first three years of taking up an appointment. And that seems to me to be something which is long overdue. The kind of training that is given is quite clearly a matter for discussion with the professionals engaged in the field. The Advisory Council has recommended that it should be a requirement, but it still has to get the approval of the Secretary of State and in the present economic conditions who knows what will happen to that. In my view it is a function of the union to see that this training is provided and that it is of a good quality.

Evaluation seems to me a quite different matter, it takes place whether teachers like it or not. There are four groups involved: teachers evaluate their own work, otherwise they would not survive within the profession. Their colleagues assess or evaluate their work, not formally, but everybody who has been in a teaching institution knows that some people are better teachers than others. Some are very good; most are considered, if you like, average, and there are a small number who are the ones who present difficulties for the unions because their colleagues think they ought not to be there. Students evaluate staff, and so do others who are interested in the institution but not part of its teaching staff. Now this happens whether there is a system of evaluation or not. It takes place in the normal course of events and if we are serious we, as teachers, have to take into account what other people think about us as well as what we think of ourselves.

I think this has a bearing on the problem of assistance. Most teaching institutions now are paying very much more attention to assisting their staff than they have done in the past. To many of us it may be far from adequate and it varies very much from place to place. But I think it is a general truth that very much more attention is being given to this and this conference is a reflection of the kind of interest that is being generated. I would like to suggest that in your discussion it is looked at very much more from the point of view of what assistance the institution can give to a new teacher or even to teachers who have been in post for some time for that matter, both from its own internal resources and by the use of outside resources. There is a lot that can be done. In-service training is still in its infancy in many areas and particularly in the field of teacher education. Some of it goes on, but there is need for very much more general discussion as to the kind of training that is desirable. We all know some of the things that are necessary: teaching techniques change and teachers need to keep up to date in their specialism. But there are many other aspects which need examination.

The second thing that I would suggest is that we ought to get away from the idea of assessment in the sense of ranking or grading. We should aim to get to the point where colleagues look at it as the normal thing to assist one another. Many years ago I was at a conference on higher education in Moscow University discussing this very problem of the quality of teaching and of improving the policy of teaching in higher education, and a professor spoke who was working for UNESCO at the time. He said that when he had been teaching it was his experience that one of the most valuable ways of improving the quality of the teaching of his staff was to go into their lectures and to talk to them afterwards. But, he added, he would never have got anywhere with it if he had not invited them into his lectures in the first instance and then discussed with them the quality of their lectures. When we get away from assessment as though it was only related to tenure and look at it as related to quality, we will be on the way to a rapid improvement to the quality of teaching all round.

Marcus Merriman

I will try to be as brief as possible. So far, I think, a great deal of the discussion has, perhaps unfortunately but nonetheless understandably, centred on the Agreement between the AUT and the Universities Authorities Panel. I think this is a bit dangerous; if university finances continue as they are at the moment, there won't be any more probationary candidates at university and thus perhaps some of the discussion on teaching will not be taken very much forward.

I think it must be recognised that, for obvious reasons, the AUT had some soul-searching over the document which you have seen, but in the end we agreed to it because it marked an enormous advance in the conditions of service for a large number of our colleagues in a large number of universities. Whereas many universities in the late 60s had evolved some procedures governing the operation of probation and its application, particularly the application of law and confirmation, the vast bulk of universities in the United Kingdom really had no procedures whatsoever. The relevant committee met and its decision was final and this made the role of local AUT committees, not to mention that of national officers who were trying to handle such cases very much more difficult than would otherwise have been the case.

And so the document that you have was brought out initially, I think, to try to rectify the situation whereby probationers could not be confirmed really by dictate of a committee without any of the guarantees in the document before you. So our first lookout was simple employment conditions, to place on the university an obligation to give proper and helpful advice which would enable a man to gain the skills necessary to hold down his job. Some universities will be wanting to emphasise that those who are responsible for teaching to teachers should know something about it. In other words, what we are trying to do through the agreement and through the Coordinating Committee for the Training of University Teachers is to train university authorities to go beyond the rather simple and somewhat crude advice which was given to young probationers. I remember myself, when I was a young probationer, that I went to a colleague and the advice I received was 'Make sure that your fly is zipped, make sure you have a glass of water with you, speak clearly, have a bit of chalk, and finish off on time'. And that was about it, as far as my lecturing training went.

Now, probation is a worrying and testing time for any new man, and I think that is one of the reasons why Dr Kennedy emphasised that those who train teachers should not include the man who ultimately makes the recommendation as to his confirmation. This is a tricky one but I think one thing that must be understood, when one is looking at whoever is to take over the responsibility of helping a probationer learn to teach, is that there is a certain amount of subservience involved. I well remember my professor, in a small department of four, three men and a boy as he called us at the time, saying 'Call me Austin, Marcus,' and I always said, 'Yes, professor.' How one gets over this problem of subservience, the problem of self confidence and the problem of help is a tricky question. I think one thing implicit in the AUT document, is that we need to ensure that there is an element of personnel management which is learned by university professors and by those who are in charge of a probationer's programmes, and I don't think that this can be overstressed. Now, of course, one thing which the document did not go into is, and which is the core of this conference, is the manner in which one assesses how well a man teaches. The problem that we found here is that most of the data which are most easily perceived about any man's teaching competence are really pretty bureaucratic. Does he get his essays marked, or in time, and returned; are his comments helpful; are his grades entered into the departmental register; how do his students turn out in degree results; are his lectures quoted back to him in essays as a subtle form of flattery; do people laugh at his lectures; do people even go to his lectures? But these are very crude pictures. Without a really detailed look at a man's teaching (and one which overcomes this problem of subservience and worry), I don't think we are going to get very far in the manner of using teaching 'quality' as an assessment, as a factor in confirming people in their probations. I think assessment, as far as I have seen it in several universities, is still clearly inadequate and what we really need is much more work in this area.

On a personal note: in the evolution of any criteria by which we judge good teaching I hope that the importance or research will not be downgraded. It is my feeling that a good researcher tends to be a good teacher. I don't want to go into that in too much detail, but because I do research in sixteenth century Franco-Scottish–English relations, I think I am a better teacher about the sixteenth century in general, and I think too much ignoring of the values of research would be detrimental to teaching. Another proboem, I think, is that we have got to recognise that learning and teaching are very idiosyncratic. My style of teaching might be totally inappropriate to certain students. One tries to get over this but one is trapped, for better or for worse, by one's personality. On the other hand, my style of teaching might work very well with certain students, and I think when we are trying to work out any norm, that one should be careful not to make everything bland.

Alastair Stewart

From my own memories as a student, and I am sure from the memories of one or two people who are at the conference as students, some lecturers don't even get the simple advice that Marcus Merriman referred to: speak clearly and have a piece of chalk somewhere more sensible than on the floor, where a colleague is trying to pick it up. The water will get knocked over, or the notes get blown over because the windows have been left open, and there is riot and revolution within the lecture theatre. You said earlier on that perhaps the Director of Middlesex Polytechnic is here as a token 'poly-person'; I suppose it could also be argued that I am here as a token student, and yet the difficulty there is that I am a token representative of a very clear majority within the teaching situation. Students within the teaching situation are the consumers, as was said in the introduction to the conference. They are critical consumers. They are investing a period of their time by entering into an institution of further or higher education to receive, to gain a body of knowledge; to receive and to gain certain tools with which to analyse that knowledge and then research for further knowledge within their particular discipline. In that respect they are critical consumers of those people whose office it is to assist them in these processes. But nevertheless that educational process is a two-way process and in terms of looking at how students behave, the academic staff are always on the top of the list for providing advice and guidance on how we should be directed and how we should be treated. However, it is only now, I think, and in perhaps recent years, that students are rising up on the list of those people who are consulted on how teachers should be assessed, on how courses should be assessed. I think that that is a very positive development and I hope that the people, whom I represent will enter into that relationship, that two-way process, positively but continue to do it critically.

The further point that I want to make is that I believe very strongly that the customers, so to speak, the consumers, should be involved in the production system, in the production process. I do think if there is a genuine positivism on both sides of the teaching process then it is the education process itself that will benefit. The students clearly will do well from such an arrangement as will the teaching staff, but the overriding factor is that both of us, and I speak collectively, have a commitment to that education process itself which is so continually under attack from all sectors, and it is the kind of attack that allows the government to reduce the amount of money it is prepared to spend on us as a sector of the community. So I think the positivism that will result from that cooperation will be to the advantage not only of both sides but to the wider concept which is the education profession. Nevertheless, it is important to point out from a student's point of view that we have certain dilemmas. A lot of students are keener to have an examination situation. A lot of students are keener just to go in there and critically over a period of two, three or four years gain that body of knowledge, get on with the job, receive the meal ticket and get the hell out of it as quickly as possible. I think that that is a very negative and unfortunate approach, but nevertheless is an approach which, I am utterly convinced, each and everyone of you has come across. So I think that we have to break down that kind of assumption that that is what being a student is all about. In just the same way as we should try and break down the assumption that being a member of an academic department means more than just being a competent researcher or that most dreadful thing, a competent administrator. So if we accept the premise that students should be involved in this process, you will have to accept, as indeed, I will have to accept: you may be involving very reactionary opinion. You may be involving a body of opinion, a students' opinion, which has not yet historically related to the need for progressive developments within the educational system. All I would say is, it is not their fault, it is the fault of the educational system, because they have been taught not to be critical throughout their primary and secondary education, and, lo and behold, they arrive in a further education or higher education institution and they are told in the prospectus 'what we hope to do in this college is to teach you all about economics and tell you how to be critical but we are not going to tell you how to be critical after you have started being critical'. Quite obviously they buck at that, because (a) it is a new development and (b) there are not the mechanisms for them to put forward those criticisms in a positive and constructive way. I think we have a two-way obligation to try and create those particular structures.

NUS policy, as so many people point out at great length, is not always the policy of every single student in the country. That, of course, is a point of debate which we are very happy to enter into with Alastair Burnett and other such people. But the policy that we do have as a union, as a national union, I think is exceptionally positive on these areas and we have just finished debating as our national conference in North Wales a lengthy, and I think, competent motion on education, policy and philosophy generally and we now have mandate as a national executive to follow up a lot of the issues that you have been talking about and will continue to talk about through this conference.

I would like to raise three final points. First, there is a need to develop a strict method in which you can assess us and in which we can assess you, and from that analysis arrive at structures and theories which will enhance the education process and break down some of the 'fence' philosophy which exists between staff and students. Secondly, I think that there should be the establishment of formal, officially recognised and officially financed, courses for both new members of staff and more experienced members of staff. The Director of Middlesex Polytechnic will be more than aware of the experience of bringing together students from a polytechnic background and students who are studying in what was a teacher training college, and I think all of us can take a leaf out of the book, which is teacher education and teacher training, and say, 'yes indeed, we must rigorously train young men and women who wish to be teachers', but let us not then assume that they have got it for life, and they are skilled for life. The concept of in-service training is one that perhaps has had not such an airing as it deserves for university, college and polytechnic lecturers. I think it is a decision that you would have to take. But apart from being educationally positive it would have a tremendous significance as far as the students were concerned. It would break down that dreadful ivory tower myth, that we are the students, and you know it all. If you were prepared to admit, as you do occasionally, academically, but also in terms of methods and procedures, that you didn't know it all and were prepared to attend a course, I think, that yet again you would positively enhance the relationship between staff and students. Thirdly, a lot of talk has gone on about the function of university and college lecturers and the debate between research and teaching. I would finish by saying this: research, in my opinion, is not teaching, but teaching of itself can be exceptionally significant research in terms of the students, the methods of teaching and the kind of problems you come up against, and I would urge you, having completed that research which is teaching, to put into practice what you find out.

Section II: The Theory and Practice of Evaluation

The Evaluation of Teaching in Higher Education

Roy Cox

The title of this paper is not precise, for it could be concerned with the assessment of lecturers, or research into different teaching methods or the evaluation of courses or educational systems. These *could* be discussed in isolation but if we are interested in evaluation primarily as a means of improving teaching and learning then all three are plainly relevant. If we are particularly concerned with finding some way of rewarding teaching so that staff will not feel they are damaging their career prospects by putting more time and effort into their teaching then perhaps we need to pay very little attention to evaluation of different teaching methods as such, but it would not be wise to ignore questions of course evaluation which put both teaching and learning within the broader context of the functioning of the department and the institution. But the evaluation of courses could itself focus on different aspects each requiring a different approach. Processes could be separated from outcomes and outcomes evaluated in terms of achievement – are the students able to do what the course was encouraging them to do – and intrinsically – is the course right in encouraging them to do it (Scriven 1967). Technical considerations may be more important in looking at performance and reflection and analysis of values more relevant to intrinsic evaluation. But empirical research into what professionals actually do in later life can of course be crucial in deciding whether a course is doing what it should be doing regardless of whether it is doing it well or badly. This raises the question of when performance evaluation is most valuable. Should all courses be evaluated during the course, at the end and some months or years later? And is meeting stated performance criteria enough to ensure professional competence?

The question of research into the functions and validity of different teaching methods are not considered here but we shall be looking at broader issues of evaluation of our courses and teaching partly for their own sake and partly to put the narrower problems of assessing teachers for probation and promotion within a broader context.

In the last few years I believe there has been a broadening of our conception of the role of the lecturer. Whereas five years ago the phrase 'assessment of lecturers' may have generated associations with students' unconstructive and rather threatening remarks about mumbling and mannerisms with a focus on lecturing, now the very title lecturer often seems inappropriate and staff are seen more as 'managers of learning' and books have even been written with titles such as "The Management of Learning" (Davies 1971). This broader conception of the lecturer's role is important in assessment of teaching whether we want to look at it simply in terms of feedback to the lecturer himself or for relevance to probation or generally for course improvement. But if we want to evaluate teaching in terms of this broader conception then the task becomes more complex and we may need to re-think not only the methods by which we approach it but also the resources of time and skill which we are willing to devote to it.

The Role of the Lecturer

This is no place for a detailed analysis of the role of the lecturer but it may be worthwhile listing some of the tasks which could be considered in teacher evaluation.

One of the most important tasks is course design. It has, nevertheless, been neglected in the past but is now given more prominence partly through the work of the Open University. A few years ago educational technologists would have been confident in recommending a detailed analysis of specific learning objectives as a first move in course design. Many would still agree but others feel that what is wanted is a concept network which gives a better idea of the structure of knowledge within the subject rather than a list of discrete objectives which can often be obscure or trivial or impossibly detailed. Nevertheless the work of Bloom and his associates in particular (1956, 1964, 1971) has been very influential in encouraging teachers at all levels to be more analytic about what they are trying to do. The hierarchical nature of Bloom's analysis in terms of knowledge, comprehension, application, analysis, synthesis and evaluation has been criticised and objectives can indeed function in a rigid prescriptive way but there is no need for this to happen if they or similar categories are taken as a starting point and are developed flexibly as an aid to understanding the actual practical teaching and learning processes. The later and less developed work on affective or attitudinal objectives has been important in drawing attention to the way in which purely cognitive objectives which dominate assessment systems can over-emphasise short term as opposed to long term educational aims.

With the exponential growth of technology, fixed bodies of knowledge and mastery of specific techniques are seen to be less appropriate but descriptions of courses which contain only lists of content areas can be of little help in encouraging students to develop styles of working and thinking and attitudes towards learning which will enable them to cope with the demands of the future emphasising adaptibility, re-learning and flexibility. Whatever our views then on the advisability of defining precise learning objectives there is plainly a need for course design to go beyond the mere listing and ordering of content areas. This is not to deny the importance of content. In the past there has not been enough justification of the selection and sequencing of content areas. This justification is perhaps as important as the understanding of skills required by the subject and is itself an integral part of that understanding. As yet it is possibly only in the Open University where questions of course design and production of teaching materials are aspects of the lecturer's competence which are given public expression. But this is a relatively objective and convenient form of evidence so it may become more prominent in the future.

But having designed his course the lecturer will then almost certainly have to actually give lectures, run seminars, discussion groups and tutorials and develop some knowledge and skill in the use of audio-visual materials and many other specialist resources, supervise individual and group dissertations and projects at possibly both undergraduate and graduate level, and develop the knowledge and ability to cope with individual student differences and learning problems – since professional counsellors cannot and perhaps should not replace the ordinary lecturer in such roles. The actual design of student assessment would come into course design but the practice of marking and criticising a student's work may still perhaps be an under-emphasised aspect of the lecturer's role. Finally, the lecturer needs to be able to asses his own courses and his own teaching. Again this might be an area where the lecturer's sensitivity to teaching problems and the way in which his course has been received by the students could be discussed and may be somewhat easier to assess than his actual teaching practice.

Criteria for Assessment of Teaching

In their very useful monograph "The Assessment of University Teaching", Barbara Falk and Kwong Lee Dow (1971) give a description of what the field of university teaching consists of and summarise this in terms of criteria as follows:

1 The objectives of a course must
(a) reflect a full knowledge of the discipline;
(b) reach the standard required to contribute to the specified degree;
(c) specify clearly the norms, knowledge and skills to be acquired;
(d) be appropriate to the previous knowledge and academic development of the students;
(e) be attainable by the students within the constraints of available time and the facilities which are provided.

2 The methods of presenting knowledge require
(a) choice of economical and efficient means for learning, whether by lectures, teaching in carrels, programmed instruction etc;
(b) other learning situations which have been devised (whether in laboratories, by presentation of essays or class papers etc) to give the opportunities for understanding of norms and for the cultivation of skills that are objectives of the course;
(c) testing and examining procedures which are both valid and reliable checks of the

whole range of course objectives. Such procedures should test the stated objectives and nothing else.

3 The teacher must communicate his interest and enthusiasm in the discipline and show a concern that individual students come to share it.

We may not wish to accept this exactly as it stands but if we approach teaching from this broad perspective it is clear that attempts to work from the characteristics of good teachers would run into difficulties. Certainly at the school level this has been the case. There has in fact been a shift away from thinking of teacher characteristics to the question of 'what is good teaching?' Falk and Dow mention that in the 68th Yearbook "Educational Evaluation; New Roles, New Means" (1969) there is no chapter on evaluation of the teacher. "The task of finding exact measures of teacher competence had proved intractable and it was seen that the problem had to be approached through the clarification of the process of teaching." We need in fact to look at the teacher's specific problems within the institutional context.

At the same time there has been a shift away from measurement of the effectiveness of teaching in terms of end results and towards emphasis on processes, towards unintended outcomes as well as those which are intended and enshrined in the objectives. There is perhaps now a greater willingness to accept the complexity of real life teaching situations and recognise a wide variety of interrelated factors, and it is difficult to assess one without understanding the others.

Models of Evaluation

This last point about complexity and interrelatedness sounds very obvious but in fact the earlier models of teaching evaluation tended to make it rather more difficult to accept the complexity as an integral part of the teaching situation and concentrated research on the more clear-cut issues. The classical model developed from work of Ralph Tyler in the 1930s has been influential. The process would centre around six stages, the formulation of objectives, their classification, their definition in behavioural terms, the specification of situations in which they might be demonstrated together with the level of achievement, the development of appraisal techniques and before and after measurements of performance to evaluate attainment of the behavioural objectives. The difficulty with this model is that it focuses too sharply upon intentions and not enough on what actually happens. Malcolm Parlett (1972) gives a very interesting account of the problems associated with this 'agricultural botany paradigm' in relation to the evaluation of innovations and higher education. He suggests a 'social anthropology paradigm' which will allow the evaluator to understand what really happens and to modify his view of what is actually important as this understanding develops. Robert Stake (1967) has developed a model which specifically takes into account the differences between intentions and what actually occurs in relation to the three main areas of antecedents, transactions and outcomes, separating off a descriptive account from the judgments.

Rationale	Intended	Observed	General Standards	Specific Judgements
	Antecedents		*Antecedents*	
	Transactions		*Transactions*	
	Outcomes		*Outcomes*	
	Description Matrix		Judgement Matrix	

Collecting information in relation to all the cells of this model would obviously be a job for the full-time researcher but for others it is a useful reminder of possible areas of ignorance which may in fact be important in evaluating any judgements or recommendations. If, for example, the observed antecedents of a course were not as intended – perhaps the previous course had plainly been a failure – then this would make the teaching on the current course much more difficult. The observed attitudes of the students on the other hand may have been quite incompatible with the type of teaching used although this may have been very well suited to the intended student attitudes. Again observation of what actually goes on in a course might reveal a great deal of independent student self-help which produces results which were not really warranted by the teaching which was actually intended. Finally a close look at the outcomes might reveal that the assessment procedures generated an approach to work which was seriously distorting of the learning and teaching which may otherwise have been perfectly satisfactory. Consideration of some of the elements of Stake's model, therefore, might be useful even if the information in relation to many of the cells is limited. Stake's model may, however, encourage under-emphasis of the wider context, but Pace (1972) gives an excellent overview of this context and Martin Trow (1970) considers evaluation in higher education from the point of view of the sociologist looking at latent functions and the institutional political context.

Methods of Evaluation

There are two main areas of evaluation requiring very different methods. One would be concerned with the processes looking at the reactions to the teaching–learning situation from the point of view of participants or observers. The other would be concerned with the outcomes measuring the changes which have occurred amongst the students attending the course. For both aspects Stake's antecedents would be relevant.

1 *Reactions to Teaching Sessions*

Perhaps the most widely used method of looking at the process of education concerns student questionnaires on lecturing. Ruth Beard in her "Teaching and Learning in Higher Education" (1972) and Dressel (1961) have discussed this and given examples and the theme has been taken up more fully in Colin Flood Page's monograph which looks particularly at the American experience of use of questionnaires in teaching evaluation (1974). The interesting thing about some more recent questionnaires is that an attempt is made to look at the responses on perhaps up to ten dimensions in relation to the student's background and preparation. Westphal working at the University of British Columbia suggests that students might rate not only their own preparation for a particular lecture but their own, say, mathematical and physics background. Other more detailed enquiries might be made into whether or not the student got lost in any particular part of the lecture asking him to write in which part it was, and which parts he found interesting or boring. Such information might encourage a lecturer to change the 'antecedents' to the lecture rather than the lecture itself.

Another way of looking at the lecturer is to use an observer who might be a colleague, a trained observer or the lecturer himself. It is possible that colleagues are very much under-used. If you ask a colleague to sit in on one of your lectures and take on the role of critic he may begin to think about teaching in new ways and question things which he himself has taken for granted for years. What is often wanted is an occasion for thinking about teaching. It is not so much that we cannot think more constructively and imaginatively about our teaching, it is just that the occasion does not arise. There is something to be said for briefing your colleague about what you are trying to do and what he might look out for – and for not doing so, but asking him afterwards what he felt you were trying to do.

In the Westphal study trained observers were used who had a check list of student responses and they would check off different aspects of the way in which the students were responding to the lecturer at different times. But perhaps more useful is audio, or preferably video, tape where this can be heard and watched either privately or in company with colleagues and students. Facilities for this are very widespread in higher education but it is not clear if they are very widely used.

But the evaluation of set piece uninterrupted lectures may become less significant as these are replaced by other methods of teaching. More recent developments in lecturing techniques which make the students much more active can generate the sort of feedback to the lecturer which can make the less integral evaluation less necessary. The type of interaction can vary enormously from students pressing buttons or showing coloured cards to indicate when they are getting lost, to answering multiple choice questions at intervals during the lecture. Or they may discuss issues or specific problems with neighbours before speaking to the whole class – this is a much more effective way of involving students than simply asking questions in the middle or at the end of a lecture. An even more structured situation has been used in postgraduate medical education (Mosel 1964). After a 20 minute lecture students spent 10 minutes writing short answers to the questions after which they went on to discuss these in groups of three to formulate difficulties for 15 minutes ending up with another 15 minutes of general discussion. This type of situation improves student learning but it can be particularly helpful to staff in evaluating their lectures from the point of view of student understanding and should reveal student interest and involvement. Such participatory lectures may not help very directly in evaluation of teaching for probation or promotion but they should enable staff to talk rather more convincingly about teaching in interviews.

Evaluation of discussion groups and seminars is much more difficult and the use of questionnaires may be less useful but Gordon Lawrence (1972) gives an example of a short questionnaire which can help to generate discussion about the quality of interaction in groups. With audio tapes and video tapes there may be more chance of interference but they can perhaps be more useful in the case of group work than with lecturing. Many universities particularly London and East Anglia have been using video tape recordings as a basis for both staff and students to learn about how to improve group work, but of course recordings can be used purely for private feedback too.

For more sophisticated group work interaction analysis has been developed, particularly by Bales (1950) and Flanders (1970). Brief descriptions of this type of analysis are given in Donald Bligh's article in "Varieties of Group Discussion in University Teaching" (1972). Such analyses are unlikely to appeal as a general method of evaluation but the categories used are interesting and their occasional use can be rewarding.

Discussion of group work by the group itself can be conducted without visual aids in the form of video recordings but there are often strong inhibitions when actually initiating such discussions. They can appear more natural, however, when there is innovation in the structuring and activity of the group – perhaps breaking a group of twelve down to smaller groups for different purposes. This generates a situation of comparative evaluation which is easier to initiate and develop.

2 *Questionnaire Evaluation of Courses*

Colin Flood Page's monograph (1974) goes into great detail about the use of questionnaires in course evaluation and the Falk/Dow monograph gives an interesting example of a particular questionnaire used in physiology. When considering the use of questionnaires it is particularly helpful to have of examples to look at and as these are widely available and will be discussed and can be seen at the Conference it is perhaps not worth going into the question in detail now, but an interesting feature of the physiology questionnaire which is reproduced in full in the monograph is the emphasis placed on evaluating students responses from different points of view. It is well known that students may see the usefulness of a particular learning experience differently if they are thinking of examinations rather than their own intellectual development. But it is very easy for this to be ignored and is then a problem in interpretation.

Similarly courses might be evaluated differently by students who spend a long time on them as opposed to those who spend very little time. Such information is invaluable in course design and yet is seldom available.

Another interesting feature of some questionnaires – particularly the student perception of teachers (SPOT) questionnaire developed by Gagné (1967) is the dual aspect of what actually happens and the 'expected ideal'. This gives perhaps a better indication of the student's satisfaction than questionnaires which ask for only a single rating. Some of the descriptions of the dimensions are interesting in their own right:

Question A: Where do you situate this course (this professor) on the evaluation scale?
Question B: Where should it (he) be situated in order that you be satisfied?

---

(46) Nature of feedback: this concerns the evaluation of the *quality* of the feedback, *without regard to its frequency*. At one extreme, there is the professor whose feedback is always perceived as depreciating the student; at the other, there is the one who always knows how to make negative as well as positive comments in a constructive way.

1 rarely constructive
2 occasionally constructive
3 often constructive
4 very often constructive
5 always constructive

Other questionnaire designs are particularly useful at revealing the different ways students might perceive objectives of different courses. Judith Stein's questionnaire used at the University of California, San Francisco, School of Medicine, is interesting in this respect. She gives ten types of course objective and asks the students to indicate to what extent they feel each was emphasised or encouraged in the course. Examples might be acquiring a body of factual information, practising basic laboratory techniques, problem solving, interpreting data or applying knowledge to the solution of problems, using and evaluating the literature or other sources of medical information. Discrepancies between different courses have sometimes corresponded to the way the staff see their courses but sometimes they have not. The actual profiles of a particular course produced by the students may be very different from that of the staff and although neither might correspond very closely to reality nevertheless the discrepancies can be a good starting point for a better understanding on the part of the student and perhaps the staff too. Information from this particular questionnaire which has other sorts of closed questions together with open questions, has been used by the chairmen of the departments to improve the teaching and learning within the school.

In general although questionnaires alone may be better than nothing, in practice it is probably best to combine sources of information about courses using interviews, informal discussions as well as opportunities for students to write more informally about their experience.

3 *The Measurement of Change*

Ultimately the success or failure of a course depends upon whether or not the students change in the desired way, but of course the relationship between this type of change and the quality of teaching is highly problematic. Notoriously large- and small-scale studies of teaching methods and teaching resources often fail to show any obvious relationship with academic performance as measured by traditional means. This is likely to be mainly due to the restricted nature of the assessments systems, Birney (1964), Abercrombie (1971), Astin & Panos (1969), but compensation for inadequate teaching must be another important factor. An added difficulty in most normal teaching situations is that there is very little pre-testing so that assumptions have to be made in order to credit students with actual gains in academic achievement. In general it would seem that pass rates and academic standards are relevant to teaching evaluation but relationship is a difficult one to interpret. At present they are perhaps of most use simply as warning signals when there are large fluctuations over time or between similar courses.

Although we might hope that more sensitive methods of academic assessment might reveal a closer relationship between teaching and achievement studies of attitude change may in fact be more promising. McKeachie suggested that attitudes can show marked change even if academic achievement does not. Attitude measurement however is particularly difficult and is not as yet highly developed in the field of education. The encyclopaedic "Handbook on Formative and Summative Evaluation of Student Leaning" by Bloom, Hastings and Madaus (1971) is one of the few books to discuss this issue and yet it may be that changes in attitudes towards both learning and the student's subject might in the long run be more important than changes in knowledge and low-level skills – although, of course, it would be unlikely that significant attitude changes would occur without some very important cognitive changes. But if we maintain a certain scepticism about the accuracy of some of our attempts to measure attitudes the actual attempt to do so can in fact be a useful way to a better understanding for the sort of attitudes we are attempting to encourage. It could be important, however, not to look upon the measurement of attitude change as a completely separate activity. The less highly structured forms of student assessment can in fact reveal a great deal about the students attitudes. Encouragement for students to keep diaries of their intellectual life and their emotional responses to their educational experience, moreover, may be particularly valuable. Although they are not as yet widely used it seems they may have an important role to play not only in enabling students to become more aware of the relationship between their education and their developing sense of identity but as a means by which staff will become clearer about the way in which their teaching relates to the more emotional and personal concerns of their students.

Conclusions

It is clear even from this rather brief look at some aspects of evaluation of teaching that a very simple approach will have serious shortcomings. The ordinary lecturer wishing to evaluate his own teaching may now see it as more threatening in that it will require more time than he had expected, so diverting him from research, and may require more knowledge and skill than he is able to bring to it. This would seem to imply that the activity itself should be rewarded as an important part of teaching competence and that time, advice and practical help should be available. Some universities and polytechnics already recognise the need for some relief of the time pressure in order to improve the standard of teaching and allow study leave for this purpose, but it may be that this will always be limited and some way of resolving the eternal conflicts between research and teaching might be attempted by enabling staff to feel strongly committed to one or the other for certain periods of time. Perhaps some subjects do require constant research commitment but others may well benefit from an 80% or 90% commitment for a year or two – although this may prove administratively difficult in small departments. But if 'experts' are used in evaluation it is important that they should not be seen as researchers who work independently and produce the answers. Co-operation is essential in evaluation studies.

Not a great deal has been written on the actual processes of appointment and promotion. Falk and Dow (1971) suggest that "attention ought to be directed to:

1 records of student progress;
2 course assessment by students;
3 any available assessments by colleagues;
4 records of any (preferably semi-structured) interview along with the judgement of experts in the field of publications and on other works which may be submitted".

They go on to suggest that detailed jobs specifications should be drawn up. Lewis Elton in his paper in this volume gives examples of more elaborate schemes which have been suggested. In America quite elaborate elections are often held and are recommended in Catlow and McGee's "The Academic Market Place" (1958). (More disturbing if perhaps fanciful pictures of academic elections are in Bernard

Malamud's novel "A New Life" and, of course, in "The Masters" by C P Snow.) There is a more promising procedure which is already in use in America and may be relevant to promotion called the 'personal growth contract'. The name is not very attractive but the idea is that every member of staff will draw up a plan or contract of what he intends to do in the coming four or five years. There is no prescription as to what should be in this contract but plainly if no mention is made of teaching developments this might be remarked upon. A discussion of the content and progress of such activities might not only provide quite useful information but be an important encouragement for educational innovations and improvements. It may well be, however, that the more significant developments in evaluation of teaching will come not from staff thinking about their own courses or from students as consumers expressing their judgements about the courses which are provided for them but by an integration of evaluation into the learning process so that an important part of the students' learning is in fact learning about his own learning, coming to understand his own interests, strengths, weaknesses, inhibitions and style of thinking and working in relation to the varieties of constraints and opportunities presented by the course. No doubt this, too, would involve formidable problems of interpretation, but just as evaluation of students' academic attainment is being more and more integrated into actual learning activities so the evaluation of teaching may develop away from retrospective and external judgements and more towards constant reflection on the significance of the educational experience.

References

M L J Abercrombie, *Aims and Techniques of Group Teaching*, 2nd edn, SRHE, London, 1971

A W Astin and R Panos, *The Educational and Vocational Development of College Students*, American Council on Education, Mass, 1969

R Beard, *Teaching and Learning in Higher Education*, Penguin Harmondsworth, 1972

R Birney, 'The Effects of Grades on Students', *Jnl Higher Education*, Feb 1964

D Bligh, 'Evaluation of Teaching in Groups by Interaction Analysis', *Varieties of Group Discussion in University Teaching*, London University Institute of Education, 1972

B S Bloom (ed), *Taxonomy of Educational Objectives I: Cognitive Domain*, David McKay, New York, 1956

B S Bloom, J T Hastings and G Madaus, *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw Hill, New York, 1971

B S Bloom, D R Krathwohl and B B Masia, *Taxonomy of Educational Objectives II: Affective Domain*, David McKay, New York, 1964

T Caplow and R McGee, *The Academic Market Place*, Basic Books, NY, 1958

I K Davies, *The Management of Learning*, McGraw Hill, London, 1971

P Dressel, *Evaluation in Higher Education*, HM Co, Boston, 1961

B Falk and K Lee Dow, *The Assessment of University Teaching*, SRHE, London, 1971

N A Flanders, *Analysing Teaching Behavior*, Addison-Wesley, Cambridge, Mass, 1970

C Flood Page, *Student Evaluation of Teaching, the American Experience*, SRHE, London, 1974

F Gagné, SPOT Test. Director, Service PERPE (SPOT), College Bois de Boulogne, 10500 Bois de Boulogne, Montreal, Canada, 1967

A P Hare, E F Borgatta and R F Bales (eds), *Small Groups – Studies in Social Interaction*, Alfred A Knopf, New York, 1965

G Lawrence, 'The Syndicate Method', *Varieties of Group Discussion in University Teaching*, University of London Institute of Education, 1972

W J McKeachie, 'Procedures and Techniques of Reaching: A survey of experimental studies', *The American College*, N Sanford (ed), Wilby, New York, 1962

J N Mosel, 'The Learning Process', *Journal of Medical Education*, 39, pp 485–496, 1964

C R Pace, *Thoughts on Evaluation in Higher Education*, Amer Coll Testing Prog, Iowa, 1972

M Parlett, 'Evaluating Innovations in Teaching', *Contemporary Problems in Higher Education*, H Butcher and E Rudd (eds), McGraw Hill, London, 1972

M Scriven, 'Methodology of Evaluation', *Perspectives of Curriculum Evaluation*, R Tyler, R Gagné and M Scriven (eds), AERE Monographs, Rand McNally, Chicago, 1967

R E Stake, 'The Countenance of Educational Evaluation', *Teachers Coll Record*, 68, 1967

M Trow, 'Methodological Problems in The Evaluation of Innovation', *The Evaluation of Instruction*, N C Wittrock and D E Wiley (eds), Holt, New York, 1970

H J Walton, 'An experimental study of different methods for teaching medical students', *Proc Roy Soc Med*, 61, pp 109–112, 1968

Evaluation of Teaching and the Institutional Context

C M L Miller

How evaluation is defined, and the means used to carry it out, depend on the answers to two questions: who is the evaluation for? and what is the evaluation for? The problems and constraints of providing evaluation data which could be formalised and used as part of the administrative procedures for probation should not lead us to forget that, first and foremost, evaluation should be for teachers and students themselves. It's purpose should be to enhance their experiences of teaching and learning; it should be useful and helpful to them. To be so, evaluation must articulate with their experiences, and must be in such a form that they can learn from it – and learn from it before it is too late.

Many staff and students have a quite understandable fear of the word 'evaluation' when applied to their efforts. Not only is there the natural suspicion that it means their being designated as 'good', or 'bad' or perhaps worse, 'indifferent', but the fear that such evaluation can only be simplistic in relation to the very complex set of phenomena that they recognise as the day-to-day reality of teaching and learning. They are sceptical that the intricate network of their experiences, which contribute to their concept of themselves as a teacher, can be encompassed in some kind of measurement which is either meaningful or helpful.

Teaching is essentially rather a private business, in the sense that its skills are so difficult to convey to others – let alone to measure. How does one describe the complexities of what makes for a good relationship with students? How does one describe the reasons behind a feeling that a course is going better than it did last year? In other words, a great deal of the educational process is intangible, subjective, changeable. A great many of the most important messages that staff communicate to their students are implicit and not obviously measurable. Not only is information being conveyed, but the teacher's attitudes, approaches and expectations. These have to be perceived or deduced by the student. A student does much more when listening to a member of staff than pick up 'content'; and the staff member does more than transmit a solid lump of 'knowledge'.

In the whole field of education it is now unfashionable to think of learning as being merely the taking in of information on an input–output model, and there is now more concentration on the *process* of learning. Surely the same kind of thinking should apply to teaching and its evaluation? It would be mistaken to evaluate teaching as if it were only about the conveying of information and the on-stage performance of the teacher. One needs to look at the whole process of teaching as it is integrated with learning; to confront the complexities involved in a changing series of experiences and relationships which are developing from day to day. What are some of these complexities that ought not to be ignored by an evaluation?

First, teaching and learning do not take place in isolation; they are very much affected by the context in which they are set. Each college, and the departments within it have their own special characteristics, personality, idiosyncracies, which are bound to exert particular influences, pressures and constraints on its members. Not only do departments vary in their more obvious aspects such as size, and the subjects taught, but also in less clear fundamentals like their atmosphere and how the power structure works. The new staff member of the department has to make sense of this, to become socialised into its subculture, form a role for himself, and try to untangle what is going on. His perception of the department and his place within it will affect very much how he behaves as a teacher.

The second complexity that an evaluation of teaching must grapple with is that teachers are individuals with different personalities and ways of coping with the departmental context they find themselves in. There is, unfortunately for evaluation purposes, no standard, ideal teacher type that people can aspire to. There are no universal rules of thumb that a teacher can follow to be assured of being a 'good' teacher – a successful strategy for one teacher may not be so for his colleague. There are no ways of behaving which will guarantee success whatever the context. In research interviews I have carried out with students, when they have mentioned to me that 'so and so is a good teacher', I have asked why. From their replies, there is clearly no stock answer; their answers are almost as varied as the personalities of the teachers concerned. But one characteristic that does recur – and it is the nearest to a universal response that has come across – is the response by students that a teacher is good 'because he takes an interest in you as a *person* . . . he wants to know how you are getting on, not just in your work, but in other things you are interested in'. Being recognised and treated as a person and not a number in a crowd was something that students said was particularly important at the beginning of their courses in helping to build up their confidence. It is interesting that this response is not concerned with the teacher as a conveyor of information, but as someone who is willing to tune into what the student has to offer.

Besides the variety in both the educational context and in the teachers within it, the third complexity for the evaluation to face is the students. Not only are students also individuals who react in different ways to the same teaching situation; they are also undergoing a process of change in themselves as they go through their course. The criteria they use to evaluate what teaching means to them at any point vary depending on the stage they are at in their own intellectual development. A first-year student's meat may be a finalist's poison. A lecture may be described as 'poor' by first-year students because it was 'not factual enough' – their expectations being that they needed a lot of facts to have 'learned something'. Later on in their course the same kind of lecture may be described as 'poor' because it is too factual for the stage they have now reached in their thinking. Also, of course, it's not a simple progression for all students between their first year and their final year; they change at different rates. This is one of the reasons why analysing student opinion about a lecture, for example, is not a simple matter. This was something that dawned on me when I interviewed students after their lectures. No two students seemed to have exactly the same impression of a lecture. They said things which could be quite different both from each other and from any impression that I, as an outside observer, had got from sitting in on the lectures. As I went on discussing students' impressions of teaching with them, it was quite clear that they did not evaluate a lecture in isolation; their impression depended on both their immediate and their long-term goals; it depended on where they were in their own personal development at the time. There can be a tendency to forget that students change in what they want from their teaching at different points in their college career. Students I have spoken to set a limited value on the use of questionnaires at the end of lectures, which they feel do not get at some of the real issues of their learning experience – of which the lecture is only one element. Staff too are often confused by the variation in the responses they get if they do try to use questionnaires at the end of their lectures; they wonder how to make changes when the students do not always seem to agree amongst themselves, or from one year to the next. It is perhaps the responses to the least important questions which produce the most consistency.

So to summarise these circumstances that an evaluation must take into consideration to take account of the complex reality of teaching and learning: (1) it has to start with the special character and constraints of an individual department, set within the personality of its college; (2) it has to recognise that there is no ideal type of teacher to provide standards of 'good' teaching which are universally applicable; and that (3) the variety among the kinds of teachers is paralleled not only by the differences between students, but is complicated by the fact that the students are changing in their needs as they go through their course.

It is possible to work out techniques and strategies which do try to look at all these parameters of the whole learning process, and which do not simply isolate measurable features and attempt to put them on a scale. It is difficult to do; it needs trained researchers and professionals; it takes time and resources. There are simply not enough of any of these commodities to go round. What I think we should aim for here is that staff and students themselves are the prime movers in their own evaluation. They should not attempt to take over techniques of the social scientist researcher which need resources and expertise to make work. I believe the most effective way that staff and students can define evaluation, in terms of these complexities of the process of teaching and learning which form their own recognisable reality, is to bring that process more out into the open for each other; to set up more opportunities for dialogue between them, to discuss their own approaches, their expectations, and what they are both bringing to the learning situation. A lot of staff time is devoted to thinking about and planning the *content* of courses – what material should be covered and what left out; how many timetabled hours should be allotted; how many staff should be involved. In this behaviour, if only implicitly, they have a model of how the student is going to react to the content and to take it in. It is assumed that the student will fit this model, or will end up by fitting it. The more courses are organised and tightly planned in advance, the more precise teachers become about what is going to be taught, the more it may seem to the student that *his* input and interests are irrelevant – the course could as well go on with someone else being there – or indeed without his being there at all. Staff should admit that they need to continually rejig and update their model of the situation in the *students'* terms. There is a need for more communication, not about content only, but about how students are experiencing their subject, how they are changing in their views about it; how the teaching is different from their expectations, and how different parts of the course do or do not relate together in their heads, rather than how they are supposed to relate. There needs to be greater opportunity for the teacher to try to see things from the students' position, and the kind of learning experience he is having; and at the same time for the student to understand more of the teacher and his approach. In other words, the process of learning should be included as part of the content of what is taught.

Making opportunities for more dialogue between staff and students about learning would recognise the concern that students expressed that a good teacher should be interested in the student 'as a person', and in how he is thinking and changing; that what he brings to the situation is important. If teachers are going to indicate to students that they want them to discuss what their learning experiences are, and in what ways they find the teaching more or less useful to them, the implication is that students matter as people, that they are important to the success of the enterprise. I'm not sure that members of staff do manage to always signal this message, even when they intend to. In practical terms, how would such dialogue between students and staff be managed? It means making opportunities for small groups of students to discuss regularly with their teacher their experiences of their subject and course in its widest context, with the teacher taking a 'non-teaching' role. I do not think that this kind of dialogue is what is usually meant by 'feedback'. The term feedback suggests a fixed model into which information is fed to correct the model and keep it to its original form and intention. The purpose of the dialogue between staff and students is that the models they both have of each other's expectations and approaches should change until they come into closer approximation.

I have been struck in my own research by the number of students at all stages in their course who have said that the research interview was the first time that they had really been able to bring out and discuss their feelings about themselves in relation to the subject and courses they were taking. They welcomed these discussions. As a researcher, I was an outsider, and I wondered how different the discussion would have been had I been one of their own teachers. I recently had the chance to find this out. Mature students doing a one-year MA at Sussex in Curriculum Development, and who are themselves teachers in a number of different higher education institutions throughout the country, have had the opportunity while doing the MA to go back to their own institutions, and set up discussions with their students both individually and in small groups. They put themselves in a listening role and not a teaching one. They found overwhelmingly that it was a very productive experience. As one said, on returning to Sussex after a fortnight's concentrated interviews and discussions with students 'I have learnt more in two weeks than I have in teaching in the college for 6 years'. It wasn't that these teachers were normally distant with their students. As they expressed it, their concerns had been with running courses – with the performance of their teaching duties. They now felt that having begun to look at the teaching situation from the learner's point of view, that their teaching would change as a result.

Perhaps I can return now to the original two questions: evaluation who for, and what for. I have suggested that evaluation should be for the benefit of staff and students to help them understand each other better. That it should not be thought of as measures tacked onto teaching procedures, but that it should be an integral part of the whole teaching and learning process if it is to be meaningful.

We cannot afford outside evaluators to monitor every course – the responsibility has to lie, and should lie, with the staff and students themselves. Staff members' models of what is going on in students' heads need to be continually adjusted by trying to understand the student's frame of reference and the models he brings to the situation. This suggests that teaching should not be evaluated in terms of better and better ways of presenting information within the teacher's framework of what he *thinks* is going on in the student's head. Teachers, whether probationers or not would be learning all the time how to adjust their own framework of presentation to the students. By dialogue between them, students are made to feel that they are not vessels into which information is pumped, but are people whose interests and contribution have an effect on moulding what is taught, and that courses do not go on irrespective of their input. It is a small comfort to a student that at the end of three years, his probationer teacher is dubbed 'not-to-be-tenured' by some means or other, when he has had to suffer him through his course. Far better both for teachers and students that they should be negotiating towards mutual understanding during that time as they go along. This would lead towards providing a context of mutual cooperation and improvement in teaching and learning. This should surely be our aim in defining what we want from evaluation.

The Quality of Course Materials

Michael Macdonald-Ross

Educationalists have always assumed that subject matter experts were subject matter experts. Consequently, when faced with poor teaching or dissatisfied students they have looked to the pedagogical strategies for relief. If only teachers knew how to define their objectives, how to select the best learning activities, how to design proper assessment. . . . Now, along with the teacher's supposed incompetence as a teacher has gone a charmingly innocent belief in his competence as an expert in his own discipline. It has seemed self-evident that a lecturer in physics was an expert in physics, at least, expert enough to teach in his speciality. These beliefs have been so strongly held, so interlocked and long-lasting that they amount to an ideology.

On the other hand, if it were shown that many lecturers in universities did not possess sufficient understanding of their own disciplines, then the nature of the game would change abruptly. No amount of tinkering with the 'content-free machine'* would have the slightest effect. Instead we should have to find ways to accelerate the academic's growth of understanding of his own discipline. Now, there is an even more serious possibility: suppose subject matter experts are not only not expert; suppose they lack those foundational skills which underlie all the academic disciplines. Suppose they do not argue rationally, do not adduce appropriate evidence, cannot distinguish conceptual problems from empirical tests and so on. In short, suppose their critical and scholastic skills are radically deficient. What then?

The question as to *where the problem lies* is in the last report an empirical question. I do not mean by that (I hasten to add) that there is any *simple* litmus paper which will decide whether teaching is good teaching, or whether a university lecturer is a scholar. No indeed, it is the very complexity of such judgements that leads people to grasp patent nostrums. I mean that in the last resort where the problem lies is a matter of fact, not a matter of belief.

Thus, if educationalists persist in assuming that all problems in the teaching/learning situation are due to poor standards of pedagogy then they are in for some rude shocks. At best, they will spend a lot of effort on irrelevant activities. If educators wish to dish out patent medicines they should remember that before you can be an effective doctor you need to be a good diagnostician.

To illustrate these ideas I discuss below three ways of evaluating course materials. The first two (survey research and the systematic approach) operate outside the subject matter. The third (content analysis) looks at the internal structure of the discourse.

---

* The systematic approach to education based on well-specified objectives.

Note: This paper is an abridged version of the paper presented at the conference. The complete version is available from the author at the Open University.

Figure 1

	Course Units Reported on															
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Reports	192	242	248	269	387	777	820	831	834	724	652	652	637	614	469	361
Interest																
Very interesting, %	109 57	163 67	169 68	183 68	229 59	191 25	160 20	192 23	355 43	159 22	189 29	154 24	384 60	313 51	155 33	110 30
Fairly interesting, %	71 37	70 29	68 27	71 26	126 33	348 45	396 48	381 46	361 43	365 50	285 44	304 47	208 33	234 38	211 45	168 47
Not very interesting, %	10 5	6 2	7 3	9 3	23 6	184 24	200 24	187 23	76 9	150 21	134 21	151 23	25 4	42 7	79 17	62 17
Not interesting at all, %		1 0	1 0	1 0	3 1	44 6	51 6	56 7	20 2	30 4	30 5	31 5	6 1	12 2	17 4	11 3
Don't know, %					2 1	2 0	4 0	2 0	1 0	4 1	4 1	3 0	1 0	1 0	1 0	2 1
No answer, %	2 1	2 1	3 1	5 2	4 1	8 1	9 1	13 2	21 3	16 2	10 2	9 1	13 2	12 2	6 1	8 2
Difficulty																
Very difficult, %	2 1	8 3	14 6	26 10	20 5	228 29	334 41	341 41	114 14	227 31	255 39	228 35	50 8	24 4	84 18	70 19
Fairly difficult, %	57 30	79 33	109 44	122 45	119 31	337 43	346 42	377 45	284 34	352 49	293 45	302 46	186 29	105 17	212 45	174 48
Just right, %	90 47	113 47	104 42	102 38	169 44	126 16	96 12	74 9	315 38	114 16	76 12	102 16	310 49	263 43	135 29	99 27
Fairly easy, %	37 19	37 15	16 6	15 6	69 18	68 9	31 4	26 3	95 11	16 2	14 2	8 1	78 12	172 28	30 6	12 3
Very easy, %	2 1	4 2	3 1	2 1	3 1	8 1	2 0	1 0	4 0	1 0	1 0		3 0	38 6	2 0	1 0
Don't know, %	2 1			1 0	2 1	2 0	3 0	3 0	3 0	4 1	3 0	6 1	1 0	4 1	1 0	2 1
No answer, %	2 1	1 0	2 1	1 0	5 1	8 1	8 1	9 1	19 2	10 1	10 2	6 1	9 1	8 1	5 1	3 1
		Socialisation					Economy and Society				Money, Wealth and Class					Spatial Aspects of Society

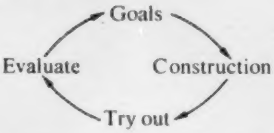
The most extensive kind of evaluation we do at the Open University is *survey research*. Since our students are at a distance there are many things we need to know about them. The annual cost of the various kinds of survey work comes to something like £100,000, so it is quite a considerable industry. If the questions you are asking are very simple (eg when did you watch the television?) you can just about interpret the responses. However, when you start asking subtler questions a whole range of problems occur. If we take Figure 1 you can see there is a strong relationship between the subject matter and the student's estimates of difficulty and interest. This is quite typical; it is not just an isolated incident. It is not clear how this kind of information is to be interpreted. As far as I know nobody inside the University has ever been prepared to say what you should do with results like this. So one of the recurring problems of surveys is their interpretation. That is a non-trivial problem.

We may also have doubts about the validity of reducing the astonishing variety of human behaviour to neat rows of figures. There is really no way a string of figures can represent the experience of individual students. That does not mean to say that you should not collect information, it simply means you always have got to think about the variety that is being lost. You may well end up throwing away a great deal more than you keep. That again is a serious limitation of survey research.

Finally, there is the question whether any kind of number collecting can be objective or ideology free. There is always a tendency for rows of figures to look value neutral, which they never are. They may even be systematically biased. In the literature on survey research there have been many experiments to show how slight changes in question format produced great changes in the responses that people make (Payne 1951). For these reasons survey research methods are best limited to the most straightforward questions. The data collected provide some weak but marginally useful indicators of course quality.

Figure 2 The Systematic Approach

The systematic approach builds in evaluation thus:



In this model evaluation becomes simply 'have the pre-set goals been achieved?' Minor adjustments to the goals are allowed; this does not alter the sharp division between 'relevant' outcomes – which you look for since they are related to the goals – and 'irrelevant' outcomes.

The systematic approach is a content-free machine; it claims to work on any goal-directed human activity. It is one kind of *rational planning*.

The second kind of evaluation in which we invest effort is the *systematic approach*. This is a form of rational planning where you say beforehand what you want the student to do and then afterwards you go along and find out whether they can do it (Figure 2).

Systematic approaches to training have been going as long as recorded history. Perhaps the first systematic approach was the systematic approach to teaching religious procedures, followed by the systematic approach to teaching rhetoric. The industrial society of the nineteenth century accelerated the growth of the systematic approach. Here Solzhenitsyn spells out one of the reasons why the German army was so effective in the early part of the First World War:

> "All the senior officers of the German army were taught a uniform system of military science. This was the legacy of the elder Moltke, who reasoned that since for a country to be endowed with a military leader of genius was a matter of luck and the fate of a nation should not be allowed to depend on chance, therefore by the application of military science it must be made possible for even mediocre men to put strategic concepts successfully into effect."
>
> *Solzhenitsyn "August 1914"*

Also from nineteenth-century Germany we have the systematic approach to the teaching of chemistry (chemical laboratories and methods of analysis). The systematic approach to education is just one of a whole myriad of ways in which people prescribe how things should be done and evaluate them by seeing whether or not the goals have been achieved.

Now, there is a lot to be said for the systematic approach. It is a way of learning: by cycling round you can improve the whole system. The systematic cycle makes it possible to discuss education inside the framework of a system. However, there are limits to its usefulness. Although you are allowed to change the goals a bit you can't change them radically. Contrast the rigidity of preset goals with the teaching process seen as an act of exploration by the teacher. At least *sometimes* teaching is an act of exploration on the part of the teacher. Maybe those are the most important occasions for that teacher. In which case, it makes little sense to ask for the detailed pre-specification of goals before you are really sure where you are going. Moreover, all kinds of things happen in a learning situation which you can't predict beforehand. Now, that does not mean to say they are not real. For example, students will often learn more from the style of teaching than they will from the actual substance of the subject-matter. They will learn more than facts and conclusions. Nevertheless, it is very much easier to spell out lower level cognitive goals beforehand. So what do you do about those subtle but real outcomes which cannot easily be pre-specified?

Does the systematic machine have any parts inside? Its purpose is to make education more effective by having well-defined goals and well-defined procedures for achieving those goals (see references in Macdonald-Ross 1973). But it may be that the machine has got no bits and pieces inside! Where the method works it is not because good methods and procedures are in-built but because particular teachers have had good intuition, they have made the right guesses and then they packaged it as though it were part of the system. My standard complaint about the systems approach is not that it is a bad idea but that people generally underestimate how much it leaves to intuition. You still leave a vast area open to the native judgement of the individual teacher. That is usually not recognised by proponents of the system.

The third method of evaluation is to look at the content of the teaching. For want of a better phrase, let us call this *content analysis*. The literature of content analysis is extensive, and I have no intention of reviewing it here, save to mention the excellent accounts by Carney (1972) and Holsti (1968). In the discussion below I shall use only those techniques of criticism and scholarship which could reasonably be expected to be the stock-in-trade of university level teaching.

Figure 3 Content Analysis (the teaching materials in Figure 3 are taken from the Open University Foundation Course in Social Science (D101))

"Government Intervention

> On this line of reasoning, unemployment is the result of a deficiency of aggregate demand in the economy. The quantities of goods that buyers want to have are not sufficient to require the employment of everyone who wants a job. To get rid of unemployment, the government must do something to increase aggregate demand. In Britain, the government is itself a main contributor to demand. It is not only a major employer of labour in its own right (in 1973, central and local government employees numbered about 1·5 million, while the armed forces were another 360,000), but it is also an important buyer of commodites that range from aircraft to school blackboards. If it steps up its own spending, this will stimulate increased employment in the industries whose products it buys. In addition, it can boost the demands of other spenders by reducing taxes: cutting the rate of income tax, for example, will leave more income in the hands of consumers, and will allow them to buy larger quantities of the whole range of commodities that come under the heading of 'consumer goods'. Still another possibility is for the government to use its control over the banking system to make more credit available to business firms, so as to encourage them to spend more on new equipment, thereby stimulating output and employment in the machine-making industries. The government's problem, indeed, is not any lack of ability to affect the amount of aggregate demand, but the difficulty of knowing just how much additional demand to create. If it goes too far, it could set up a demand for labour which is so great that it exceeds the supply; the demand for output will have risen to a point at which it is not possible to satisfy it, and the result will be a general rise in prices: it will have got rid of unemployment only to find itself faced with a problem of inflation.

(Figure 3 Cont.)

All the same, these difficulties do not invalidate the general principle that the government has the power, by manipulating aggregate demand, to prevent the kind of unemployment that occurred in the nineteen-thirties. It will be shown presently that this is not, in fact, the only kind of unemployment that can occur: but it is undoubtedly the most important kind to get rid of. When the demand for labour throughout the economy falls short of the supply, no conceivable effort on the part of the unemployed themselves can make any difference to the situation. No matter how hard they look for work, it is not there to be had; however willing they are to move to other areas, or to start afresh in new industries, it is simply not possible for them all to find employment. It is to eliminate this kind of heart-breaking situation that the government should take action to stimulate demand."

The Opposing View (which students are not allowed to see. Patrick Hutber (*Sunday Telegraph*, 23rd February 1975) prefaces this section with a comment that "all Governments, of whatever persuasion, are committed to full employment . . .

Faced with this threat of unemployment, Governments pump out money to "stimulate demand" and reduce the unemployment. In the short-term it works, or appears to work, but its longer-term effect is to prepare the way for another rise in prices and another rise in the underlying rate of unemployment.

Each time a larger stimulus to demand has to be given; each time it has a weaker effect. Each time the adverse consequences are more serious.

Thus the monetarists do not say: create unemployment. They say, do not over-expand the money supply in a futile attempt to get it down. You may, it is true, have to sit through an uncomfortably high short-term rate of unemployment to stop the long-term increase in inflation and unemployment.

In the meantime, there are many steps you can take by way of removing economic rigidities and by way of retraining to bring down the long-term unemployment rate. You can and should be generous to the unemployed though you must realistically recognise that this in itself may tend to increase the length of time for which they stay out of work."

"The most humiliating fact about our present discontents is the contrast we present to those countries which we either defeated or liberated during the course of the last war. They prosper and flourish, despite the increase in oil prices, the cost of raw materials and inflation generally, while we bicker among ourselves, lag behind in production, excite mingled derision and contempt from our critics, and arouse pity from those who would wish to remain our friends.

> But the same contrast establishes one conclusion beyond a shadow of doubt. The cause of our troubles is not economic, nor has it anything to do with world conditions or our loss of empire, or with any of the other easy excuses which we are so ready to accept, and our present rulers to invent, for our comfort. It is a disease of the spirit for which there is no one to blame but ourselves.
>
> There is no objective factor in our present situation with which our neighbours have not had to contend since 1945, except those which, at the outset, rendered their situation far less promising and more difficult than our own. There is no possible excuse for our bad performance, which is due, and in my opinion entirely due, to facts wholly within our own control as a nation.
>
> It is, in short, a failure of the determination, of the will, and even of the desire to think out problems out in a rational and unsentimental fashion. The fault lies not in our stars, but in ourselves that men do thus abuse us.
>
> Obviously everyone who has lived in this country as an adult during the past 25 years must bear some share of the blame for this national decline. . . .
>
> But the greatest influence of all, at least in my view, is what used to be called the *trahison des clercs*, the moral betrayal by those who ought to be upholders of the traditional virtues, qualities and institutions, and of the truths and values which it was their duty to maintain.
>
> Incidentally, it is never the proletariat, the so-called under-privileged, or the victims of alleged social injustice who destroy a society. It is always a failure of moral leadership, and intellectual integrity, or, worse still, an element of conscious betrayal, on the part of those whose function it is to counsel, to bear responsibility, and to guide, that causes moral disintegration of the sort which I am seeking to describe.
>
> The best and the most faithful of these must at least bear the shame of failure. The others must, in greater or less degree, carry the responsibility for actively bringing down the edifice their forbears erected, and which they inherited as a sacred trust.
>
> The French Revolution was long preceded by a betrayal of this nature, and so, I think, was the Bolshevik triumph of 1917. (Lord Hailsham, *Sunday Telegraph*, 23rd March 1975.)

We start on Figure 3 under the heading 'Government Intervention'. Here is quite a nice piece of prose; it is reasonably coherent and well argued. The rest of this teaching material has a similar pattern. Now, the snag is not what the author has done so much as what he has not done. He hasn't told the student that the whole theoretical basis for this is Keynesian economics, and that the action programme is a Keynesian action programme. And he does not let unfortunate counter-examples like stagflation (which is not supposed to occur in a Keynesian economy) raise obstacles. He does not let the student know that there are radically different views about what you should do when unemployment occurs. The excerpt from Patrick Hutber is one such example. It does not matter which of them is right, if either. What matters is that the author has tailor-made the teaching to present one point of view only where other reputable and interesting views exist.

The excerpt from Lord Hailsham is a further shock. The little piece from Hutber was an economic argument but now Hailsham introduces a moral argument. You may ask whether it is fair to suggest that an economist should concern himself with moral questions. But the course team deliberately choose 'real life' problems so as to show how the procedures of the social sciences were superior to common-sense thinking! Having chosen such interdisciplinary problems the team were surely bound to treat them in all their complexity. What is the consequence of the systematic bias of this text for the students! One consequence might be that he never discovers the strength of the alternative view. But a more serious consequence might be that he gets the message that systematic bias is acceptable reasoning.

Figure 4

to demonstrate the inadequacies in *much* everday thinking	*Most problems can be evaded if you insert an escape clause*
As you have worked through the *various* discussions, you will have discovered (even from the small sample of analyses presented here) that there are various theories and *emphases* concerning the causes of crime	***Discussions, analyses, theories, emphases . . . are they so vague as to be interchangeable?***
trying to keep total unemployment to a minimum, it should also *take what steps it can* to see that its average duration is as short as possible, and *pay attention* to the needs of special groups such as older workers and those whose skills are not likely to be in demand again	*You can see clearly how vagueness is used as a defensive manoeuvre*

Now consider Figure 4. The comments made here are micro-criticisms. When I first started annotating this text I thought these were all local errors. As time went on I began to see that they are systematically connected. In these extracts there are many vague terms. Nobody is going to disagree with phrases like 'The government should take what steps it can' (that is like being in favour of motherhood). But, by the same token, such a vague statement is virtually tantamount to saying nothing. The classic criticism of vagueness in teaching is that vague language heralds vague thinking (ie teachers who exhibit undue vagueness in their courses are people who have not really thought their subject matter through). But is there more to it than that?

In Figure 5 you can see vagueness connected with what Thouless used to call

Figure 5

[1] *From the community's point of view*, it is [2] *obviously wasteful* to have a large number of its workers doing nothing; total output is reduced by the amount of goods they could have produced if they had been in jobs	1 *What does this vague phrase mean?* 2 *This is a 'dirty trick' which begs various questions, for instance, does* $GNP \propto$ *employment?*
In this context some would regard the different approaches as mutually contradictory – if one is right, the rest must be wrong – and certainly the tone of some writing on this general subject leads fairly easily to this conclusion. A number of other social scientists, however, conclude that these differing approaches are complementary rather than contradictory: first, since the term 'crime' covers a number of *different* kinds of activities and means different things in different contexts, different causes may well be in play in different cases; and secondly, several causal factors may be operating simultaneously in any one case, rather than just one.	*Altogether a more subtle case* *How might such a question be settled?*
In coming to a preliminary conclusion for your own interest (as well as for your TMA this week) you may wish to choose one of these three lines of approach (represented diagrammatically below) or some variation on them, *or* you may wish to take some completely different line. This is for you to decide	*Apparently, it's just a matter of taste!*
... factors which, as we have seen, is also involved. We have to take account of the wider, economic, administrative, political, social and psychological *factors*, beyond the single individual. *Whether he knows it or not**, the individual is likely to be *affected by* these *factors* which are largely external to him, and it is thus essential for the social scientist to study them	* *Invalidation of purposive explanation*

'dishonest tricks in argument'. You can see in the first example 'from the community's point of view' (What the phrase means is pretty unclear); and 'it is obviously wasteful'. The word 'obviously' is a giveaway. It begs all kinds of questions. It is not entirely clear that gross national product varies according to unemployment. There are all kinds of sophisticated issues which are being systematically dodged here.

The next example in Figure 5 says, in effect, 'Some people think that these approaches are complementary and some people think they are diametrically opposed'. That is all very well, but you notice the author does not tell you how to settle a question like that. And then you get the suspicion towards the end that it is just a matter of taste. At first sight you might think, 'Well, that may be a bit sloppy, but it's not a fundamental issue', but as we go on I hope to show you that these are very serious issues indeed.

The last example in Figure 5 again has vague words like 'factors' and 'unaffected'. To say something affects something else is not to say anything very precise. Then, all of a sudden, you find the strange phrase 'Whether he knows it or not'. Now, it seems to me to be quite interesting that this phrase comes in where it does. The crux of the author's intention is to avoid the notion of personal culpability (whether or not the person who commits a criminal act is responsible personally for his actions). The author wants to avoid that question. And you notice that the phrase 'Whether he knows it or not' slips in, in a very sly kind of way. The effect of that phrase is to invalidate purposive explanations of behaviour, and therefore rule culpability out of

Figure 6

Becker states that "a useful theory of criminal behaviour can *dispense with* special theories of anomie, psychological inadequacies, or inheritance of special traits and simply extend the economist's usual analysis of choice" and that the "general criterion of social loss is shown to *incorporate as special cases, valid under special assumptions,* the criteria of vengeance, deterrence, compensation, and rehabilitation that historically have figured so prominently in practice and criminological literature." (1968)	*Subsumption* not *rejection*
Becker is saying that for a broad range of criminals everything we have been told – that they are sick, abnormal, deviant, or deprived – is *wrong*. Criminals are relatively simple, normal people like the rest of us, and any attempt to treat them as abnormal or deviant or to 'rehabilitate' them is doomed to fail.	*No, he's not saying that*

Crime, as has been argued above, should be seen as involving an audience of social actors rather than as a single behaviour with an imputable cause. Just as an explanation of how Open University students learn would be incomplete without taking account of how full-time and part-time staff organise their knowledge and teach them, so crime is inadequately explained if one ignores all those involved in its definition and processing.	*False dichotomy* *False analogy*
The main point that emerges, both there and in many discussions of the topic, is the *relativity* of the concept of crime. What is counted as crime varies from time to time in the same culture, between different cultures, between different subcultures and groups even in the same society, and also according to the circumstances and people involved. Once one considers the facts on this it becomes obvious that there is, after all, no absolute class of actions or 'typical' kind of behaviour which constitute 'crime'. What is counted as 'crime' is *relative* to a particular period, culture or situation. (This general point is elaborated further in radio programme 2.)	*Acts are behaviours which meet certain criteria. If those criteria are man-made rules, as laws are, then clearly they will not be absolute or universal. However, that does not make the concept of crime unclear; nor does it allow for acts to be* ambiguous *in specific times and places.*
Crime is 'created' by society, then in the sense that the criminal is one who has been so labelled	*This is the conclusion the author has been determined to reach . . .*
Crime, then, is a quality conferred on some forms of behaviour by an audience; it is the latter that eventually determined whether any item of behaviour is criminal or not. The study of crime becomes inseparable from the study of social control agencies like the police, courts of law, and legislative institutions, since crime consists in a transaction between rule-breakers and social control personnel	*Recall, as you read this, that the author hopes to 'demonstrate the inadequacies in everyday thinking' and 'introduce the methods whereby scholars . . . etc. etc.'*

court. This led me to think there are systematic troubles underlying these apparently superficial items of vagueness.

Now consider the first example in Figure 6. What Becker actually says is that he can subsume other theories of criminality under an economic theory. Then the commentary says that Becker holds that other theories of criminality are wrong. Now, that looks at first like a harmless error in scholarship. The second writer has not noticed that Becker holds a subsumption theory. But, as a matter of fact, it seems essential for the second author's argument to reject and to discredit Becker. And the first steps to discrediting somebody is to state his argument in a form which is clearly ridiculous.

The next example shows a false dichotomy (are there no other plausible definitions of crime?). This is followed by a false analogy; there is really no logical connection at all between the two cases. We begin to see that the errors in reasoning are closely connected with the case that the author is presenting. You can read the cards how they fall. The way I read them is that the author decides on the conclusions in advance, thus has a strong view about the way he wants to go in the argument. The author seems willing to commit any kind of reasoning fallacy, providing it gets him to the conclusion. Now, that has all kinds of educational implications. I ask again, *what is the consequence of this for the student?* It is a rather troubling matter. If all students learnt were the conclusions at the end you might say, well, it is only a matter of whether or not the author is right or wrong, and maybe the author is right. But unfortunately what is at stake is the style and manner of the teaching process. It is a question of what students pick up from *the way in which the teaching is done*. If students think that is good reasoning about the subject matter, then they are learning the very things that they should not be learning. Thus, whether or not the author is right is very much secondary to whether or not the reasoning processes have been honestly and fairly displayed so that the students' mental growth is enhanced.

At the bottom of Figure 6 is the conclusion the author has been heading toward: "Crime is 'created' by society" and "Crime is a quality conferred on some forms behaviour by an audience". This is very interesting, because this particular block of our material is supposed to demonstrate to the students the inadequacy of everyday thinking and introduce the methods whereby scholars avoid these inadequacies in their professional work. No doubt you will draw your own conclusions as to whether this scholarship has demonstrated the superiority of social science methods.

Now, I must also face the question – it is just because I have chosen a social science unit on a controversial subject matter that fallacies in reasoning and conclusion hunting should be so obvious? The answer is that the criticism about teachers hunting conclusions is a pretty old one in the literature of education. Joe Schwab (1962, 1969) made it about teaching in science fifteen years ago (and, no doubt, Dewey before him) and it seems to be still valid. So I think that it is more than just a little local difficulty.

The issue we must think about is whether or not this kind of critical content analysis (perhaps more organised, more constructive and better integrated with the teaching process) could be useful as an evaluation method. I am inclined to think that it is essential if we want to make some kind of thorough going improvement in the quality of teaching. I think it is likely to be threatening although I think it could be done with more subtlety than I have showed here. It is very likely, however, that very few organisations or individuals would have the tenacity to carry a programme like this right the way through their teaching work. If they did, I think the effect would be very great indeed.

I think it is worth asking ourselves what we want out of the evaluation of teaching. Do we want just low level ranking (whether students prefer this teacher to that) or do we want a thorough going nuts and bolts, bone and muscle attack on the substance of the teaching? I think the answer to the latter question is 'yes'. The doubt I have is whether any of our institutions at present could contemplate doing this on a large scale. Two steps would be necessary, both of which cost money and human resources. First, some of the teaching must be recorded in permanent form so it can be studied at leisure. Second, a group of scholars and educational advisors must get together to do the analysis. In my opinion the benefits of content analysis (or should I say critical scholarship?) are worth the serious attention of modern universities.

References

T F Carney, 'Content Analysis: a Technique for Systematic Inference from Communications', Batsford, 1972

O R Holsti, 'Content Analysis' in Lindzey and Aronson (eds), *The Handbook of Social Psychology Vol II*, Addison-Wesley, 1968

M Macdonald-Ross, 'Behavioural Objectives: a Critical Review', *Instructional Science 2*, 1, pp 1–52, 1973

S L Payne, 'The Art of Asking Questions', Princeton University Press, 1951

J J Schwab, 'The Teaching of Science as Enquiry' in *The Teaching of Science*, Schwab and Brandwein, Harvard University Press, 1962

J J Schwab, 'College Curriculum and Student Protest', University of Chicago Press, 1969

c

Is it Possible to Assess Teaching?

L R B Elton

Well, throughout today, that question surely must have been in the minds of most of us. If you expect me to answer it you may be in for a disappointment. Let me say that I want to start with asking the question 'Is it possible to assess anything?'. I then want to go on to a subject that has already been touched on today, it is a relationship of research and teaching, and discuss briefly some attempts to assess teaching and the attitudes of university staff to this exercise. This will bring us to the question of the difficulties that are in the way of this process, and I finally want to suggest that some help in answering the question which is the title of my talk may be obtained from teacher training.

Is it possible to assess anything? Let me tell you that I have never yet been asked to conduct an opera at Covent Garden. They say that I am not good enough, but I have failed to get from them a list of criteria let alone behavioural objectives which would tell me that I was good enough to conduct the opera. Let me give you another example. A David Hockney today costs considerably more per square inch than an Alma Tadema. Does that mean he is a better painter? An Alma Tadema used to cost a lot more per square inch than it costs today. Does that mean that since his death he has become a worse painter? Again, I am sometimes asked by another university whether x is worthy of promotion, say, to a personal Chair. How do I judge this? I put it to you that I judge this by comparison with others whom I know and who either do or do not hold Chairs. In fact in all the cases that I have discussed we do not have absolute criteria, but we compare one professional with others in the same profession and put them on an ordinal scale. So what we assess, I maintain, is not in any absolute sense here but in a relative sense comparing performances by different people. Now that, I believe, is the clue as to why we never seem to question whether we can assess the research achievement of an academic staff because we know lots of his colleagues in similar fields all over the world perhaps, and we can match him up against them in our minds. That, I believe, is basically how we judge a person's research achievement. Certainly it would not be possible to obtain criteria that would satisfy anyone regarding what is a certain level of research achievement. How much research is enough for a lecturer, how much for a professor? As Cleopatra said to Anthony: 'If it be love indeed tell me how much,' and Anthony answered: 'There's beggary in the love that can be reckoned.' That, I think, is a quotation that should be engravened on our hearts when we have to discuss the kind of problem we are discussing at the moment.

Let me turn to research and teaching, and I am fully aware that Colin Flood Page is in the audience and, indeed, I got one in on him in advance because as you will see the first reference on my list of references is to the article to which he was referring earlier today. The question is frequently asked whether there is a positive relationship between good teaching and good research. It's a question to which, as we have already discovered, the academic profession tends to give a rather unthinking *yes* as an answer. Even Halsey and Trow in their monumental work on the British academics, my reference 2, seem to take it for granted. Their only reference to this problem of the relationship between research and teaching is the rhetorical question which they ask in connection with the low rating given by engineers in their sample to research, and the question they ask is this: 'How does one teach the branches of professional engineering and applied science without engaging in research in those fields?' That is all. The assumption is immediate – you cannot. The evidence is non-existent. My interest in this question is not, however, in the answer, whether there is a positive relationship or not, but in the fact that the majority of academics believe there to be a positive relationship, that they also believe that it is possible to assess good research and yet that it is not possible to assess teaching. How one can then establish a positive relationship or any kind of relationship between two things, one of which everybody agrees can be assessed and the other one the same people agree cannot be assessed, I cannot imagine. On the other hand, 76% of the sample thought that promotion was 'too dependent on published work and too little on devotion to teaching', and interestingly, when asked whether they consider their own department to be above average, average, or below average in a particular area of activity, while the majority of academics considered their own department to be above average in all its activities, this was particularly so in its undergraduate teaching. How they judge that since apparently it is not possible to judge teaching, I do not know. Perhaps you are beginning to realise that I have some doubts as to how my academic colleagues use their well trained minds when it comes to attacking this thorny question of problems in teaching.

Let me now turn to attempts at assessing teaching. Well, we have already had some talk about the assessment by students in terms of questionnaires and I may say that this is done mainly of lectures. I have come across little, if any, of student assessment of teaching activities other than lecturing. You will hear more about that from Colin Flood Page who has written what I am sure for a long time is going to be the standard monograph on the subject, but let me give you a quotation from a much earlier day. Max Planck as a student is talking about the lectures which he received from two of his Professors, Helmholtz and Kirchhoff: "I must confess that I did not profit significantly from my lectures. Helmholtz was obviously never properly prepared, he spoke haltingly while looking for any data, he needed, in a small notebook. In addition, he constantly made arithmetical errors on the blackboard and we felt that the lecture bored him at least as much as it did us. In contrast, Kirchhoff lectured from carefully prepared notes in which each sentence carried its proper weight and was in the right place. Not a word too many, not a word too few. But the whole gave the impression of having been learnt by heart, dry and monotonous. We admired the speaker but not what he said." I think this is an immensely perceptive analysis of two kinds of bad lecturing; let me also add that it is the sort of thing you would expect from Max Planck, who indeed in despair decided to read original papers, read Clausius on energy and lo and behold before we knew where we were had discovered quantum theory. But of course our students in the majority of cases are not Max Plancks. So we may get less perceptive reactions from our students than this one, which I am sure Planck was much too diplomatic to confide to anyone except his diary, but nevertheless we ought to listen to them. Now Flood Page in fact puts the matter very nicely: "If teachers think of it in terms of absolute student judgement on their worth as people, they are not viewing the process in proper perspective. If, on the other hand, they are clear about what effect they want to have on students in a classroom, those students are the best instruments for measuring that effect." Let me add here that anyone who has taught both in school and in university knows that in school he gets feedback from his class, which makes it unnecessary, generally, to pass round questionnaires. He does not get that in a lecture theatre, and when a teacher – and lots of people have often quoted this remark – says he has taught for twenty years, he often means that he has taught for one year repeated twenty times; because he has not been able to profit from the feedback of the students as it wasn't there. Now in fact Bradford University has been the one that has done most along the lines of getting towards student assessment of lecturers. Smithers at Bradford concluded that rating scales would be a technically sound method to assess lectures, but nevertheless Bradford University rejected the method in spite of some persuasive work for it by Foy, another member of the University. Both these are quoted in my references. I would think that the prejudice against using student questionnaires to help one improve one's teaching must be overcome, but I would like to stress that there are serious limitations to what you obtain from such questionnaires, not only because of the clearly distorted feedback that comes back, but much more seriously because it is virtually impossible to apply this method to any method of teaching other than the lecture. And the sooner we get away from thinking of university teaching as lecturing and nothing but lecturing, the better.

Let me now turn to more general assessment. Most of the work here has been done actually in Australia, and the monograph by Barbara Falk and Kwong Lee Dow explains what is done there very well. I found an interesting quotation in an American report "Faculty Development in time of Retrenchment": "Teaching, unlike research and publishing, remains very much a private professional act, rarely open to collegial scrutiny. Effective teaching remains the step-child in the hierarchy of academic values. It could be greatly improved by opening up the process to sensible and substantive criticism." Here we hear the crucial point, namely that teaching must become more public, although some of us may feel after Macdonald-Ross's talk that we would like to hide ours under a stone. But it won't do. Now in this country in the last few years there have been at least two attempts at listing a number of criteria to be looked at in terms of criteria for teaching in connection with promotion. There may have been more, but I will refer to my own some four years ago at the request of the Association of University Teachers, and more recently the one in the Nuffield yellow book. In mine, which is represented in Figure 1, and I think I should add that the article in the AUT journal was headed "A Personal

Figure 1 Criteria for promotion (Elton)

1 scrutiny of course and project work of students
2 willingness to experiment and innovate
3 involvement in curriculum reform
4 prepared teaching materials
5 use of student feedback for improvement
6 knowledge of new developments
7 attendance at meetings and courses

View", so that it does not commit the Association. I suggested that there are a number of things one can look at in what a teacher does that would give clues regarding his teaching. I would exclude examination results, which I would imagine most of us would agree with, although this has said a lot about what our examinations are like. To scrutinise course and project work and even the lecture notebooks of students would be a useful exercise. How many ever, when considering the appointment or promotion of a university teacher, look at the theses of his students. An interesting exercise. Similarly, we could do this at undergraduate level. Next, the willingness to experiment and innovate in one's teaching. So often I am asked 'how often can I improve my lecturing' when perhaps the answer should be 'should you be lecturing'. Then there is involvement in curriculum reform, the preparation of teaching materials which can be looked at, the use of student feedback for improvement – does one have the courage and perception that one can learn from one's students? – knowledge of new developments elsewhere, attendance at meetings such as this.

The Nuffield criteria, see Figure 2, are at an altogether higher level. They relate to the sort of thing you might do in teaching, which is really very close to what you would be judged on in research; publications etc. contributions to conferences. I would therefore suggest that my criteria might be considered as part of what one might look at in getting a probationer towards tenure or maybe promotion across the bar, while the Nuffield criteria are much more like the ones one would look at in promotion to a

Figure 2 Criteria for promotion (Nuffield)

1 publications about teaching problems
2 published or publicly available course materials
3 participation in inter-university projects
4 contributions to conferences
5 supervision of theses on educational topics
6 consultancies on teaching issues
7 use of external referees

readership or promotion even to a Chair. I am not suggesting these are the only criteria. I do suggest, however, that attendance by academic staff in the lecture or classroom of the person to be scrutinised is a method that must be used with the greatest caution, remembering the uncertainty principle and its effect in measurement. Even introducing a camera has quite serious effects on the activities of a teacher. However, there is a totally different method which we can still use and that is of self-assessment. It is perhaps not unreasonable that a person who is supposed to be as reflective in his attitude to evidence and all this kind of thing as a university teacher, should be able to say something himself about what he is doing, in other words give an account of himself. I owe to Bob Woodbury, a questionnaire which is used in the University of Massachusetts, see Figure 3. This lists a number of particulars which the faculty member completes every year. He then hands it over to a personnel committee and head of department who make additional comments, which he reads, he then signs them as having read them and off it goes to the dean.

Figure 3 Annual Staff Evaluation Report (Univ of Mass)

1 personal particulars	completed by staff member
2 teaching activities	
3 professional work	
4 service and administration	
5 other professional activities	
6 comments of department personnel committee	read by staff member
7 comments of head of department	
8 signature of staff member	
9 comments by dean	

Now we are concerned particularly with item 2 – teaching activities – and what it is that he is asked to say something about is shown in Figure 4 where he is asked to list the courses which he is giving and which he gave in the last year, any major changes in teaching approach or responsibilities during that year, any individual supervision of students in theses and projects, any more personal responsibilities in advice and teaching, and the last one which is largely irrelevant in this country, major teaching activities in special programmes – the University of Massachusetts, like many American universities, has special programmes of a very unusual nature like the "University Without Walls" and things like that. When we have got anything as revolutionary as that, we will add this to that questionnaire. Now the point about such a questionnaire is that it wonderfully sharpens one's mind to one's own activities if one is asked to give an account of oneself, whether it is used in any kind of promotion procedure or not. If any of you have ever used the method of writing a reference for somebody by asking him to write it for you, you will know what I mean. It is a very, very good exercise.

Now let me turn to attitudes of university staff. A recent survey at Swansea, which will be reported on later, indicates considerable dissatisfaction with promotion policy. Even in research, too much weight is given to quantity. Let me add that since most of our research is mediocre a lot of mediocre work is better than a little of mediocre work. They want a point system for teaching, research and administration with clearly defined criteria. I am in favour of some kind of a point system so as to give weightings, which the AUT also proposed some years ago, but I don't think clearly defined criteria are easy to obtain. Could I ask, however, whether these authors would accept students on promotion committees? That of course, conflicts with the reserved areas that we hear often about. I suggested this in my AUT article and perhaps that was one of the reasons why it was labelled a personal view, but I do think we ought quite seriously to consider whether students have not a place on promotion committees if we intend to use teaching as a serious criterion for promotion. In contrast to the view put forward by the Swansea group, which I think is very widespread, let me however present you with an opposite view. Lipson, for whom I have the greatest regard, is indeed a first class teacher himself, in fact so good a teacher that he doesn't realise how difficult teaching is. "There is much talk these days of the relative importance of teaching and research in universities. The rank and file think that research is overrated and that the good teacher is as good as a good researcher. They would like to see teaching given as much weight as research in considering candidates for promotion. I do not agree with this view. Teaching is our basic work for which we should not expect any extra reward." Let us just be clear what the opposition is because it is very well expressed in that view.

Figure 4 Teaching activities

a courses
b major changes in teaching approach or responsibilities during the year
c supervision of thesis and project students
d responsibilities in advice and tutoring
e major teaching activities in special programmes

Let me turn to the difficulties which face us in assessing teaching. The difficulty is that the product that we are assessing, the educated student, is a remove from us, unlike in research, where we assess directly through the research paper, which is the product. Secondly, there is a long tradition of privacy in the process of teaching, and thirdly, the relationship between the process and the product is not well understood. In other words, we do not know what kind of product our teaching will produce, there is no simple relationship. Judging teaching by judging the finished student, even if we could do it, would be a difficult and chancy undertaking, but it also means that since this is what we are after all basically interested in, namely, the quality of the finished student, it is extremely difficult to know what is good teaching.

Finally, however, I would suggest that a possible answer to the question of my talk lies in teacher training. Here I would like to say that we do seem to have put the cart before the horse. Would we not consider ourselves a little foolish if we spent long hours discussing how to examine our students, but nobody had either taught them or decided how to teach them or even knows how to teach them. We surely would first of all learn something about how to teach our students and then decide how to assess whether they have learnt what we have taught them. The processes of assessment then would become clarified through the activity of teaching. I cannot see why this should be different in the training of teachers and I therefore suggest that the reason why we cannot find sensible and simple answers to the question 'can we assess teaching?' is that we haven't got any answers yet to the question 'how can we train university teachers?'. If we had substantial training programmes, at least one would hope that the trainers would become better teachers since we all learn most by teaching. It would certainly clarify our aims of the teaching process. We would build up a body of knowledge which would tell us what to look for in a good teacher. We might, in the process of developing training methods, identify good teachers as those who are acceptable as trainers and the process of training should develop a questioning and creative attitude to teaching. It is this which is perhaps the most important. I detected throughout the day a certain amount of prejudice in the remarks that were being made. In other words, remarks like 'it is my belief' or 'I think' etc, not 'the evidence for it is so-and-so'. Remarks which academics regularly make in connection with teaching and rarely in connection with research. If we could but develop a questioning, a creative attitude, to teaching in universities in the way we develop it in research, I think many of the questions we are asking at the moment would get answers and answers that are backed by evidence. So I have no easy answer to the question which is the title of this talk 'Can we assess teaching?', but what I would say is that, until we have developed substantial training programmes for university teachers, we cannot expect a sensible answer to that question. I therefore close by pleading for the extension of well researched training programmes and it is the 'well researched' I would like to stress. If we developed well researched training programmes for teachers, then I believe, many of the answers regarding how to assess teaching will fall out.

References

1 C Flood Page, "Teaching and research – happy symbiosis or hidden warfare?", *Univ. Quarterly*, 27, 102, 1972

2 A H Halsey and M A Trow, "The British Academics", Faber and Faber, (a) p 322 (b) p 350 (c) p 356, 1971

3 C Flood Page, "Student evaluation of teaching – the American experience", SRHE, 1974

4 A G Smithers, "What do students expect of lectures?", *Univ. Quarterly*, 24, 330, 1970

5 J M Foy, "A note of lecturer evaluation by students", *Univ. Quarterly*, 23, 345, 1969

6 Barbara Falk and Kwong Lee Dow, "The assessment of university teaching", SRHE, 1971

7 T Hore, "Promotion through teaching: some issues and problems", *Notes on higher education*, Higher Education Research Unit, Monash University, November 1973

8 Group for human development in higher education, "Faculty development in a time of retrenchment", *Change Magazine*, 1974

9 L R B Elton, "Teaching criteria for promotion", *AUT Bulletin*, p 6, June 1971

10 Group for Research and Innovation in Higher Education, Newsletter No 4, p 4, Nuffield Foundation, 1974

11 R Startup and M M Gruneberg, "Time to look afresh at ways of getting ahead", *Times Higher Ed. Sup.* 7.3.75

12 H Lipson, "Teaching ability", *Physics Bulletin 25*, 379, 1974

Student Evaluation of Teaching

Colin Flood Page

To begin with, might I just refer to the question put by the last speaker, the engineer, who asked how could he convince his colleagues of the value of evaluation? I think it worth calling attention to the fact that he is on an easier wicket in an engineering department than many others. Our experience in introducing innovations in teaching, evaluation of teaching, and all this kind of thing, is that departments of applied science are more ready to listen than most other departments perhaps because they are practical people, and can see more immediately the practical benefits of what is being suggested. People who really run a long way away from such things most of the time are people in arts and humanities departments, and it's worth pondering why this should be so. (It is only an aside, but an aside worth keeping in mind: the longer I study higher education the more I find that these differences between subject fields are extremely important in all kinds of ways, some unexpected.) Perhaps also our engineer colleague had at the back of his mind that well known quotation from Samuel Butler (the seventeenth-century one): "He who's convinced against his will is of his own opinion still." It's certainly often in my mind when talking to colleagues about these matters.

In starting to say something about evaluation of teaching in the United States, I would like first to explain how I got into this, because I think it is important that you should understand what the roots of my own interest are. When I moved to Bradford from London four and a half years ago, I arrived at a time when proposals were being put forward that student evaluation of teaching in the University should be used as a major factor in consideration for promotion. In a rather scattered and desultory fashion I had already come across a lot of references to this topic in the States. I had begun to realise that it was quite a complex operation, and so I decided that I would look further into it, rather than rest content with simplistic opinions and uninformed emotional reactions. In other words, my interest in this grew from a very practical consideration. When I begin to look into the American work on this subject I was quite surprised, and other people have been since, by the sheer quantity that exists. In my monograph I believe I have used something like 170 references. If they are not enough for anybody I can provide another 200 or so, not yet used. In other words, it has been a topic of very great interest; and writing of various kinds, ranging from meticulously reported research to extremely polemical and passionate exhortations, abounds in American books and periodicals. I think it is important that those of us who speak American as well as English should try to translate this material for the benefit of English teachers, because, as in many fields of educational research where much work has been done in the States, we should see what use we can make of it, and not merely start blindly and ignorantly off on our own, supposing that nothing else has been done and that we are venturing into a new field where all is untried and all is unknown. This does not mean, of course, that we should not take note of the fact that there are differences, sometimes quite considerable differences, between the American situation and our own. For instance, there is a difference in the atmosphere, the ethos of education, particularly higher education. There is, I would say, a more commercial orientation towards education in the States: a basic idea about accountability and value for money which is only now beginning to be taken note of in any serious fashion in this country. The whole business of evaluating teachers and teaching is part of this general idea of accountability.

My brief is to talk about evaluation of teaching, and what I shall be speaking about is mostly student evaluation thereof; because that is what has been more written about and more researched into in the States. For those interested in looking at the general picture of how evaluation of teaching is done in the States, what sort of measures are used, who does the evaluation and what sort of use is made of it, a good guide is Lee's "Improving College Teaching".[1] There have been other surveys since, but I think that one is in some ways the best. What came out of that general survey, in line with one or two previous ones, was that the most usual method for evaluating teaching was some form of hearsay, ie the head of a department, or someone like that, picks up on the grapevine general ideas about what a teacher is like from his colleagues and from odd students whom he happens to talk to. In one way or another that always seems to have been the most common American form of evaluation of teaching. I think that you would agree that that is also the way it works in this country.

This suggests an interesting line of thought. We heard one or two remarks yesterday about the question of the legal aspects of evaluating teachers and teaching. To me it is a very intriguing development, because the legal way of looking at things and dealing with things is so very different from the academic way. In this particular instance hearsay is regarded as valid evidence in important decisions about teachers. Any lawyer, of course, would put that out of court at once. He would say that hearsay is something on which you should never found any major, large-scale process if justice was to be done; yet we accept that it is normal and proper. In fact, we tend to run away from any more systematic, open kind of procedure. The general idea of informal, personal interaction with students is something which we consider to be an essential part of the education process. We think it is a good thing; and a reasonable method which we are quite content to use in the business of promotion or assigning worth to teachers. Therefore there is something worth a lot of thought in the question 'do we want to make evaluation of teachers more accountable in the American sense, more legal in the sense that has been raised by some people yesterday?' We are asking about something which is uncomfortably close to the heart and nerves of many of us.

Now, as to the question of student evaluation of teaching. It is very difficult to reckon how many institutions of higher education there are in the United States because of the vast spectrum ranging from the very high grade, respected university, to extremely doubtful organisations. I suppose one might say there are something like 1,500 to 2,000, perhaps more. Of these (although it is difficult to get very exact figures) at least several hundred, maybe well over a thousand, are using student evaluation in one form or another, though not necessarily as part of a promotion assessment programme, as I shall explain later. It is therefore a widespread phenomenon; and it takes usually one of two major forms, of which one is more common than the other. The chief form is to have a questionnaire of some kind (which is probably standardised within a given institution) making certain specific enquiries about the teacher's performance. These forms are given out at set intervals, filled in by students, collected, and either processed by the teacher himself or by some central office. That is the more usual method. The questionnaires themselves vary quite a lot. Some of them are very brief and have been very carefully researched, with different items carefully graded and those that were not very good removed. Others are fairly lengthy. Some are really rather sketchy affairs. Thus while the overall procedure is more or less standardised, there is still scope for quite a number of variations on the main theme. The other kind of evaluation, less frequently used, is undertaken at the end of a course or when people have got their degrees. The graduates are asked for a sort of retrospective essay, once they have passed through the process. They are asked to look back at what was done to them, and to say what was good or bad about it. I haven't come across many accounts of this technique, but it does exist here and there.

How has this developed historically? It appears to have been going on in the United States in one fairly organised form or another for about 50 years. The University of Washington is usually credited with the earliest development of systematic evaluation by students, almost exactly 50 years ago. There has also been other work going on, nearly as long as that. (It is interesting to note that if one looks back to writings of the twenties and thirties very much the same kind of comment, very much the same kind of weighing up of the operation, could be found then as had been found since.) It went on in differing degrees with varying attention to proper scientific detail for quite a number of years, not on any very widespread scale, until maybe the late fifties and sixties; and then students in particular began to get interested in it. The American counterpart of the NUS in 1949 recommended that students should start evaluating teachers, but it was some time later before it began to develop on any large scale. Come the early 1960s, before the more political troubles of the late 1960s, this kind of operation became very widespread; but much of it was transitory. Quite a number of institutions took it up and dropped it again. In many cases the short life was due to the fact that the process was very much student inspired and, as one or two speakers were indicating yesterday, when students are invited to take part in any kind of formal university process or generate it themselves, they don't tend to continue with it for long if no immediate and obvious results are seen. So there was a tendency for it to be dropped. Then there was a lull, with continuation here and there. Subsequently staff have tended to take it over much more than students. Some teachers, who were perhaps rather afraid to begin with, found that there was a certain value, a certain use for it, and therefore they themselves took up the running. For instance, a recent article in the Journal of Chemical Education[2] reports a survey of 200 chemistry teachers in the States on this topic: 129 replied, and of these 112 (87%) had used student ratings regularly and systematically. Their value was more problematic. Thus, though it is difficult to get clear figures, it does appear that the general picture is that student evaluation is fairly widespread in the United States and has persisted for quite a long time. One therefore supposes that it has a certain usefulness; but, whether this usefulness is tied in closely to certain, very specific American features rather than to features that we can translate over here is, I think, a point that needs considering.

The use to which evaluations are put is not standardised. If we speak first of all from the point of view of the students: a lot of student unions and other organisations of that kind have produced little booklets in which they list evaluations of staff, comprising brief descriptions of their courses, as well as some kind of numerical listing of their good and bad points. They frequently ask that a member of staff who has himself been evaluated should write something on how he sees the situation. In a very large university (and some of the American ones, as you probably know, run to an enormous size), this kind of thing can be most useful. A detailed, graphic and lively picture of what actually goes on in a course, instead of merely what it says in the official booklet, can be very helpful, particularly in the large commuter universities where the unofficial grapevine hardly exists; and 9 am to 5 pm students don't really know much of what goes on. (It is an interesting fact that when one looks at some of these booklets from institutons where there is a sharp division between the science and arts faculties, such a division may be reflected by two widely differing styles in the booklets.) That is then one major use, by students.

From the point of view of staff, of course, there are two really separate operations, which have been touched on several times in this conference. First of all, the individual staff member himself can use evaluations for his own improvement. In many institutions the rule is that if a staff member asks for this evaluation, it is processed by an evaluation office, but the information derived from the evaluation is not made available to anybody else unless the staff member so wishes. That is one side of the coin. It is simply a question of using some kind of instrument to measure what the staff member is doing and the staff member himself saying 'well, yes, from this I have gathered so and so'. He can either make use of it or not, as he pleases; but it is entirely for his own private consumption. Hopefully, it will help him to improve if improvement is necessary.

On the other side, of course, there is the whole elaborate apparatus of using student evaluation as part of the documentation for staff promotion. Here there are varying degrees of publicity given to the material. In some institutions the material is available only to the individual member of staff and to the chairman or the head of department, who may or may not take it further. He may just use it as material for discussion with the staff member, or he may formally submit it on, eg an annual basis, as part of an exercise designed ultimately to weigh in the balance for promotion, tenure, etc. It may go still further: it may be the rule that all evaluations of this kind are passed on to the dean or some kind of higher administrative official. It is easy to see, therefore, that there is a good deal of variety in the use made of these things and a good deal of difference in the importance given to them, the degree to which the administration takes notice of them, etc. Certainly there is much activity, though difficult to pin down precisely. In this country I am aware of quite a lot more student evaluation going on than is actually recorded in the literature; but I don't think it is anything like as widespread as it is in the States. There are many questions that need to be asked about the process, and in our present context, perhaps the most important one is: how relevant is it to our situation, as distinct from the American one?

References

1 C B T Lee (ed), "Improving College Teaching", Washington, American Council on Education, 1967

2 M E Schaff and B R Siebring, "What Do Chemistry Professors Think About Evaluation of Instruction?", *Journal of Chemical Education*, 51 (3), pp 152–154, 1974

The Politics of Teacher Evaluation and Improvement: A Look at the United States

Robert L Woodbury

Recent attention in Britain to issues of teacher evaluation and improvement in higher education seems unlikely to abate in the near future. Since a concern with the assessment of teaching has been carried further in the United States, it would seem potentially helpful to review what is currently being done, how this intense concern with teacher evaluation developed in the States, and of what significance any of this activity might have for higher education generally. I speak not from the perspective of an expert in the field but as a full time academic administrator of a large university in the US who has long been buffetted around, politically and academically, by the issue.

The evaluation and assessment of teaching, and concern for its improvement, is supported at my university and many others with an elaborate paraphernalia of procedures and reports, practices and programmes, questionnaires and forms. A not atypical pattern of devices directed at the evaluation and assessment of teaching includes the following: (1) systematic provision for the mandatory assessment of all teachers in all courses through student questionnaires two times each semester. In some institutions the results are publicly available; in others the data remains confidential. On my campus, teachers and departments may employ a standard questionnaire adopted by the university or use an individually prepared alternative that a particular department considers more appropriate to its particular teaching approaches or nature of the discipline. (2) All academic staff on our campus complete an annual report describing their own professional activities. (Naturally, there are white, green, yellow and orange copies for each level of the academic bureaucracy.) This report allows faculty to summarise their activities in areas of research and scholarship, administration and community service, and various areas of teaching. The sections dealing with teaching request comments on one's courses, graduate and undergraduate, guidance activities of theses or independent study, changes one has adopted in his or her curricular or teaching approaches, teaching plans for the future, and involvement in teaching in 'special' programmes such as interdisciplinary seminars, residential courses, in off-campus settings, or to atypical student populations. These reports are reviewed and commented upon by the department head, personnel committee, dean and occasionally by the provost. (3) All major personnel actions – for tenure, promotion, major salary adjustment – require, at our institution, documentation of the staff member's accomplishments and qualities as a teacher. This documentation might include results of student questionnaires, written assessments by colleagues, even letters solicited from a teacher's former postgraduate students. The nature of the evidence – or at least the quantity – has begun to parallel that we formerly associated with the evaluation of a staff member's research and scholarship. (4) The full force of the moral persuasion of the system in emphasising the importance of good teaching. This is most clearly signalled by the Board of Trustees of an institution like the University of Massachusetts officially indicating that the first general criteria for tenure and promotion is 'a record of outstanding teaching'.

But the formal and direct instruments of review and assessment of university teaching are also supported by a variety of other programmes and procedures that elaborate a concern with good teaching and interact with devices for assessment. Let me suggest just a few: annual salary adjustments as a reward explicitly for good teaching; 'teacher growth grants', analogous to research grants, to support a teacher's development of skills or new curricula over the summer holidays; teacher improvement centres (on our campus there are two, often in competition); students as regular voting members of curricular and personnel committees lending a 'teaching' or 'consumer' bias to the deliberations; and a variety of other inducements and stimuli directed toward a more conscious and explicit support for the improvement of university teaching.

This elaborate, sometimes almost ludicrous, concern for the quality of university teaching and mechanisms for assessing it was almost unknown a decade ago. On my campus most academic staff believed that one's publication record was 'the name of the game' and whatever lip service was paid to the importance of teaching carried little weight in the world of reward and punishment. The belief was probably well founded. How, then, did this quite remarkable change in North American universities come about? What accounts for a shift in emphasis and concern that is reflected in the paraphernalia of approaches to the evaluation and improvement of university teaching that I have attempted to outline? Although the story is a complex one, I would suggest several areas or aspects of an explanation that may be of interest to those viewing similar developments in Britain.

First, and possibly most important, was the role of undergraduate students in the 1960s. The upsurge of student power, in both highly publicised forms and in quieter ways, had an incalculable impact upon curriculum, teacher attitudes, educational methods and many university policies. Not surprisingly the impact upon teaching and learning issues was far greater than upon issues of research. Significant efforts at teacher evaluation and assessment began at many places, moreover, at the instigation and often under the exclusive control of students. But most importantly the very demonstration of student power – or the perception of it by staff and administration – often stimulated a concern (some might say fear) about problems and possibilities in university teaching.

Second, there has been over the last decade steadily increasing public or political concern about issues of teaching and learning in higher education as the enterprise has become increasingly larger and costlier. Some have seen these pressures in the form of demands for 'accountability'. But whatever the appropriate term, the effect has been one of increasingly searching, even strident, inquiries as to the quality and efficiency of teaching in higher education. An extreme expression of this public concern was marked in the state of Massachusetts last year when the state legislature passed a law mandating faculty teaching loads in universities. This I hope is an extreme example. But the more general pressure has resulted in greater university attention, often in public and formalised ways, to issues of teaching quality. British institutions can expect a similar pressure to continue to develop, particularly in hard times.

The third area or arena I would cite as providing a context, if not a direct cause, for attention to teacher evaluation was the climate of innovation and change, alternatives and reform, that characterised the rhetoric and even the behaviour of a decade. A climate of educational critique and innovation had two results probably more important than the reforms themselves: first, proposals for change stimulated debates all over university campuses about issues of teaching and learning; and, second, teachers found themselves teaching in uncomfortable or unfamiliar situations where they were forced to reconsider approaches to their own subject and their own teaching. An historian lecturing to 200 white middle-class nineteen-year-olds on the industrial revolution found his teaching (and often his research) dramatically unsettled by interaction with 'university without walls' students who were older, less white, and had jobs and families. And so issues of teaching and learning, and assessment and evaluation, were as much the consequence of innovation as the motivator for it.

Fourth, many campuses found that one unit or faculty or school anticipated the developments of the decade, as 'trendy' as some of them may have been, and placed the remainder of the departments in a defensive position. At our institution a controversial and innovative school of education placed students on personnel committees, required teacher evaluation by students, set up clinics for aiding university teachers, and experimented with a variety of curricular innovations before other departments had contemplated such steps. Given the climate of the late 1960s, few departments could resist such measures particularly when other departments had implemented them and university administrators encouraged them. So a psychology of 'keeping up with the Joneses' developed around teacher assessment and other related areas.

Fifth, younger academic staff found themselves in a position to force consideration of issues of teaching emphasis and quality in a way that their powerlessness had precluded in the past. Young academic staff, often on probation, have been saying for years that 'the university doesn't really take teaching seriously; all they care about is research and scholarship'. This familiar refrain had a different impact in the last ten or fifteen years. Enormous growth in higher education in the States brought enormous growth in academic staffs as well. (Our institution, for example, grew from about 400 full time academic staff in 1960 to over 1,300 in 1970.) Many were new and younger faculty who quite simply forced more democratic governance in their departments and had the power of numbers to force a shift in the importance attached to problems of teaching and learning relative to research and scholarship.

A sixth factor was the growth of the teacher improvement industry itself, particularly as it emerged at the university level. There were simply more articles, more speeches, more conferences, more discussions, and more professionals dealing publicly with the evaluation and improvement of university teaching. This heightened activity not only brought attention to the subject, but also began to suggest to regular academic staff in a variety of departments that the issue of evaluation and improvement was not all ridiculous nonsense. There was a case to be made and a place to start. Those staff who had always earned nodding approval in common rooms for sage comments like 'if students evaluate teaching, it will just encourage popularity contests that have nothing to do with rigor' or 'everyone knows that active research is necessary for good teaching' found themselves increasingly subject to more complex propositions and conflicting evidence.

Finally, it seems to me that some of the paraphernalia for teacher evaluation and assessment at an institution like mine and many others in the United States was a consequence of growth itself. As the institution and its bureaucracy became larger there seemed to be an internal dynamic that necessitated documentation and procedural regularisation (it even sounds bureaucratic!) in order to translate vague generalisations of trustees and chancellors into something operational. Further, and I think concurrent with our growth and not necessarily caused by it, was the increasingly legalistic nature and style of our personnel procedures. As an aside, I would guess that the actual legal costs of our institution have increased ten-fold in the last few years. But the main point is that the entire personnel process – tenure, probation, reappointment, salary, personal grievances, professional responsibilities – are increasingly adjudicated via the procedures, rules of evidence, form of advocacy, and guidelines of the American judicial process. (How right de Tocqueville was about the United States!) This development conspired with the growth of bureaucracy to make the methods and approaches to teacher assessment steadily more formal, written, systematic and public.

These seven areas do not 'explain' the apparently rapid shift of attention in the United States to matters of evaluation and improvement of university teaching. The story is certainly more complex. But I think they do identify some of the aspects and parameters of the story. And furthermore they may suggest what elements may or may not be present in Britain. But having said all this, what does it all add up to? What does the experience of a decade with steadily more emphasis on issues and mechanisms of teacher evaluation and improvement at an institution like mine tell us? To the extent that some control over the future is an option, what should British academics most concern themselves with in approaching the politics of teacher evaluation and assessment? Three closely related observations seem appropriate.

First, it strikes me that we should remember that the *mechanisms* of teacher evaluation and improvement are not our main concern. Our real goal should be to create an institutional *environment* where the concern for and discussion of and reflection about good teaching is a central, conscious, and public part of the university mission and not a periferal and private one. In that context formal attention to matters of teacher evaluation and assessment has been an important part of the creation of that environment. The dialogue between staff and students and among staff about issues of teaching and learning has been notably improved by, among other things, such apparently overformalised mechanisms.

Second, the creation of the environment I have described depends on a multiplicity of factors of which procedures for teacher assessment is only one. Any reasonable strategy must attend to the reward system, the power system, the role of students, resources for development, the climate for curricular innovation, and several other areas as well as evaluation. Over and over again in British universities and polytechnics, and American institutions as well, I have seen a significant allocation of resources to one or two of these factors designed to improve the quality of teaching and total inattention to the others; the result, of course, is not only a poor use of resources but a tendency to scapegoat the allocation when the overall strategy was at fault.

My third generalisation I offer with genuine trepidation. This does not derive from my lack of conviction in the conclusion but my fear that it will be misunderstood. Quite simply, I believe that it is extremely important for universities and polytechnics to adopt a careful, comprehensive system of teacher evaluation and assessment – but then not take any statistical results concerning individual staff very seriously. We should remember, as the cost benefit professionals have not, that we are talking about the quality of an environment, one where good teaching is respected and thought about and worked at; such matters must be formerly and systematically attended to. But some numbers on a questionnaire summary can tell us only a small part of what a person in a university is contributing to the academic community.

The Evaluation of Teaching – Some Comments

Alex Main

I know that some of the other speakers at this conference will have dealt with the more technical aspects of the evaluation of teaching, and so I shall be content with some few comments on the political aspects of the process.

During the last two years I have been very impressed by the efforts which I have seen the British universities make on behalf of their new lecturers. Every new university teacher now has the opportunity to attend some brief form of induction to his institution, and some elementary form of introduction to university teaching. Although we still have not become very systematic or 'professional' in this provision, we have come a long way since 1970.

Yet, despite all the efforts made by individual universities, the one development that I believe has done more for new lecturers is the recent AUT/UAP agreement on probation. Perhaps I should correct myself and say that it has the *potential* to do more than anything else for them – because I detect that universities are not being very positive in their implementation of the agreement.

The agreement puts a responsibility upon both the university and the new faculty member. On the former it places the responsibility of providing guidance, supervision and training during the early years, and on the latter that of getting the best which he can out of his university's efforts on his behalf. It is, to my mind, a charter which will do credit to the profession of university teaching.

There is, however, one weak feature of the agreement: it does not give very effective guidance to universities about the evaluation of new staff during the probationary period – except in terms of the best timing for evaluation.

University staff are chary about evaluation, especially since the ill-fated PIB pronouncements about student-based evaluation for merit awards and promotion. The result is that many of the universities I have visited since last October have made no attempts to relate evaluation to the probationary period.

I do not think we should shy away from evaluation, if we think very carefully of the help which a new teacher needs. He needs, first of all, feedback on how his teaching is perceived by his students and how it affects their learning. Secondly, he needs to indulge in a certain amount of self-examination to assess whether over his first year or two he comes up to his own standards; and lastly, he must be given some idea of how he rates as a member of the profession he has aspired to join.

My own North American experience convinces me that many of the amateurish attempts at student evaluation of teaching have not clearly distinguished between two aims: that of helping a teacher to improve his teaching, and that of assessing him for tenure and promotion. Whether the evaluation instruments have been designed by faculty or by students, they usually have no clearly stated purpose, and are usually responded to by students in a very off-hand way because students see no practical relationship between the evaluative process and the improvement of teaching. Indeed, in several North American universities, students are beginning to withdraw from the evaluation of teaching because deans and chairmen of departments use student ratings too unthinkingly to justify merit increases or tenure decisions.

I believe that the student evaluation of teaching should be kept firmly in its place: as a feedback instrument to the teacher, to assist him view the effectiveness of his teaching. That I believe to be the right of every class of students, and one which no teacher should deny.

On the other hand, I believe that students have no right to judge whether a university teacher is worthy of the confirmation of his lectureship or promotion. It is the right of every profession – and indeed it is part of the definition of a profession – that it is self-policing. The profession of university teaching is a complex one, and it demands many skills; and it is only those who are already experienced in the profession who can judge whether a novice has proved his potential to become a master. It is to that end that I have always urged that we develop a proper system of peer evaluation of teaching and tutoring skills to accompany those we already have for research and administration.

My greatest hope for the new probation agreement, however, is that it will foster the most important evaluative process of the three which I mentioned – that of *self-evaluation*. The agreement allows for the appointment of a senior colleague to give help and guidance to the new faculty member. Most departments have told me that they don't know who to appoint, nor what that person should offer. I believe that the best he can offer is to assist his younger colleague to draw up, in his first year, a personal development plan. This plan could indicate what sort of teaching, research, administrative and other experiences he would wish to have; what skills in each of these areas he already feels he possesses and which ones he would try to develop; what personal aims he wishes to fulfil during his probationary period; and how each of these could be monitored during the period.

It would then be a co-operative task for the new and the more experienced teacher together to monitor and modify the learning experience of the new teacher and continually assess progress. However inadequate any peer evaluation of his career development might be, every probationer would have at least three years of training in self-evaluation, a skill which will be invaluable in later years of teaching or in any other profession.

I believe probation must be used positively and that it must be a learning experience for the new teacher and all who come in contact with him professionally. Since we have an agreement between the interested parties which makes this possible, we should act upon it steadily and wisely.

Section III: Case Studies

Student Perceptions of BSc Prelims at Sussex

Dr Michael Eraut

This evaluation study of the BSc Prelim Course in two science schools at the University of Sussex was planned and conducted by a team of four people: Dr Michael Eraut and Mr Haydn Mathias of the Centre for Educational Technology, Dr Peter Unsworth of the School of Mathematical and Physical Sciences (MAPS) and Dr John Stamper of the School of Molecular Sciences (MOLS). It was intended as a pilot study which would open up areas for further investigation and entirely financed out of existing resources. Nevertheless it took at least half the time of member (HM) over a 2 year period, and an approximately equivalent contribution from the other three together. As I warned the conference organiser the analysis of the work is still in progress, so I cannot yet give the full picture.

The background to the study is important as it affected our aims and our methodology. In MAPS a group of Physics Tutors were meeting regularly to discuss teaching problems, both prior to and then in association with an inter-university investigation into Higher Education Learning Problems in Physics (the HELP Project). Three themes of special interest to the Sussex tutors were small group teaching, independent learning and student motivation. In MOLS a new core curriculum had just been designed and was about to be implemented. It involved integrating Organic, Inorganic, Physical and Theoretical Chemistry, and changing the teaching methods to reduce the number of lectures and increase the use of books. An earlier innovation, the BSc by Thesis Course, was being evaluated; and there was interest in evaluating the new core curriculum with a view to making further improvements. The Centre for Educational Technology had been associated with these activities in both MAPS and MOLS; and as Acting Director I was becoming increasingly anxious that any support for faculty in developing innovations in teaching and learning should not be based on an inadequate diagnosis of the problem. It was important to find out not only what features of the degree course seemed most significant to faculty, or to CET, but also which features seemed most significant to the students. Our agenda of innovations might well have little correspondence with the students' agenda of issues.

Our study had a twofold aim: general course monitoring with special attention to some recent innovations; and pilot research to try and discover what students consider to be the most significant aspects of their degree course experience. Though the study is attempting to follow a contrast of students through all three years of their course, the greatest effort has been put into the first two terms of the course, officially designated as BSc Prelims and examined at the end of the second term. The data reported below only relates to the prelims period.

We wanted the study to be intensive and therefore chose a small sample of 40 students selected randomly from 4 groups on the basis of intended majors. One transferred from Maths to Physics and one left in the first week of term so we ended up with:

9 Mathematics Majors in MAPS
11 Physics Majors in MAPS
10 Chemistry Majors in MOLS
9 Non-Chemistry Majors in MOLS.

We used three main methods of inquiry:

1 *Study Sheets* which indicate the allocation of time to various types of academic work. These were completed for three weeks, one half way through Term 1, one near the end of Term 1 and one half way through Term 2.

2 *Questionnaires* which were completed in the first week of Term 1, at the end of Term 1 and at the beginning of Term 3.

3 *Interviews* at the beginning of Term 1, the end of Term 1 and the beginning of Term 3.

Students brought questionnaires to the interviews where some of the points were amplified and additional issues raised. The first two interviews had fairly definite agendas but contained many open-ended questions. The third interview only had a map of issues which the interviewer wished to explore at some stage, and the distribution of time between them followed the student's interests. At all times the interviewers were looking for unanticipated points which might be raised or hinted at.

Our results to date will be reported under four main headings:

1 School experience and initial outlook as revealed by the first questionnaire and first interview.

2 Student reactions to prelims as revealed by the second and third questionnaires.

3 Student allocation of time as revealed by the study timetables.

4 Student perceptions of prelims as revealed by the second and third interviews (preliminary analysis only).

As yet we have had little time for investigating systematic relationships between these four batches of data by looking at individual profiles, but we hope to undertake this type of analysis later.

1 School Experience and Initial Outlook

54% of the students in the sample were from grammar schools and 20% from comprehensive schools; and a further 10% came from technical colleges. All the

MOLS students, half the Physics majors and 2 of the mathematicians did Maths, Physics and Chemistry 'A' levels (including 2 who did Physical Science instead of Physics and Chemistry). Five Physics majors and 3 Maths majors did 2 Maths subjects and Physics, and 3 Maths majors did Maths and Economics. Four students in the sample also took 'A' level General Studies. The average grades were usually B or C; and the only noticeable disparity in grades was in Mathematics where Maths majors averaged B and the other groups averaged D. Not one student in the sample did Biology at 'A' level, an interesting comment in the constraints of the sixth form curriculum. It would appear that 'modern curricula' are still having a relatively small impact on 'A' level courses. Only a quarter of the sample had done SMP Maths, a sixth Nuffield Physics and an eighth Nuffield Chemistry.

The nature of 'A' level work was of particular interest to us as it could be seen as an indicator of being prepared for university. Twenty-one students stated that they did mainly set work and 13 that they mainly worked independently, with 5 in a mid-way position. When questioned about which style they preferred, most students opted for what they were used to, though there was a detectable shift towards a preference for independent learning and 18 preferred this mode of study, 15 preferred set work and 6 remained neutral. Teacher encouragement of independent work was 'high' for 12 students, 'fair' for 15 and 'low' for 12. Perhaps it is not surprising in the light of these differing experiences and preferences that the time spent on independent study varied enormously, ranging from 2 to 18 hours a week.

The type of set work most frequently encountered was 'Problems or Exercises' (34 students) and this was followed by 'Essays' (16), 'Reading' (13) and 'Past "A" Level Papers' (11). The preferences followed a fairly similar pattern. It was also significant that, when questioned about their areas of greatest achievement, 20 attributed their success to their teachers and only 8 to their own efforts, 5 suggesting a combination of both.

At the beginning of their interviews students were asked how they thought being at university might be different from school or college; 26 mentioned increased responsibility for their own study, 13 personal freedom and independence and 6 living away from home. No other response was all that common. When asked more specifically about study habits, 21 thought they would need to change and 18 that they would not. The pattern of response related quite closely to their school experience, as 14 of the 21 whose school work was mainly set by the teacher thought they would need to change their study habits, but only 2 of the 13 whose school work was mainly independent thought they would need to change. When questioned about the type of change which they thought might be necessary, 11 mentioned increased responsibility for study, 10 self-discipline to work, 8 more time spent on study, 6 more private reading and 7 adapting to lectures. When asked about their relationships with tutors, 22 thought it would be different from that with their teachers at school; and 16 of these 22 thought that their tutors would be more personal and informal or more helpful. Given the favourable attitude towards school teachers mentioned above this gives tutors a lot to live up to!

The major source of motivation in the sixth form appears to have been to get to university (20 students), followed by a strong interest and satisfaction in the work (9 students). When asked what features of their 'A' level courses they particularly liked 24 students mentioned specific subjects, 8 mentioned aspects of the teaching–learning process and 6 said 'nothing'. A surprising number disliked the 'A' level courses in their own subjects: 4 out of the 9 maths majors, and 2 each of the physics and chemistry majors, while 11 of the 30 non-mathematicians disliked mathematics. When asked about areas of success over half the students gave their intended major subjects, but only 10 were prepared to attribute their success to enjoyment, interest or satisfaction.

The choice of subject at 'A' level seems to have been primarily determined by a mixture of interest and achievement, with the influence of teachers being the only other factor of significance, and when it came to choosing a major subject at university, interest, achievement and a wish to study the subject further dominated all other considerations. Reasons for entering university were, however, much more diverse. 'A strong interest in studying a particular subject' and 'to broaden my outlook' were the two most important reasons, and these were closely followed by 'a wish for freedom and independence to develop as a person', 'a chance to participate in the social life of a university community' and 'a chance to participate in an intellectual community where issues and problems are freely discussed'.

2 Student Reactions to Prelims as Revealed by Questionnaire

The second and third questionnaires were almost identical in order to facilitate comparison of student reactions to Term 1 and Term 2. Three main components were included:

1 Five point scales on various aspects of lectures, tutorials, books and the course as a whole
2 Open-ended requests for comments on the above, and for notes on the best and worst features
3 Topic charts indicating the novelty and difficulty of individual topics within each course

Although the information reported below is highly specific to the Sussex context, I have included considerable detail for two reasons. First, because many of the problems revealed are in my view typical of first year science courses in higher education; and secondly, because it indicates what kind of evidence can be obtained by questionnaire. (This will be contrasted later with the kind of evidence obtained from interviews.) It should, however, be stressed that our questionnaires were followed up by interviews and all interviewees were promised copies of interim reports and a chance to discuss them. So they were more likely to be taken seriously than in those situations where the issuing of 'feedback' questionnaires is perceived as the normal routine.

Evidence was gathered on four main courses: Structure and Properties of Matter (SPM), Maths 2A (Calculus and Vectors), Maths 4A (Analysis and Linear Algebra), and Equilibrium, Structure and Bonding (Chemistry 3A). Each student took three courses, selected according to the following pattern:

	SPM	Maths 2A	Maths 4A	Chemistry
Mathematics majors	Some	√	√	X
Physics majors	√	√	Some	X
Chemistry majors	√	√	X	√
Non-Chemistry majors	√	√	X	√

Some mathematics majors took Economics as the third course, and some physics majors took Principles and Perspectives of Science. One physics major took an electronics course.

2.1 *Variations Over Time, Between Courses and Between Student Groups*

Students were asked to provide overall ratings of the *courses* on a number of dimensions. The dimensions rated and a summary of the results for the Autumn term are as follows:

Presentation – Average to Good
Problem Sheets – Slightly more difficult for the MOLS students but generally at the right level
Linkages with other branches of the subject – A little weak in Maths but otherwise good
Aims – Fairly clear in all courses except Maths 4A
Interest – Above average for most groups and up to 4·7 for Maths majors taking SPM. Maths 4A is surprisingly low for Maths Majors (2·3); MOLS students much preferred their Spectroscopy to the Thermodynamics.

Results for the Spring Term show that only the Chemistry course maintained a steady rating. All the other courses decline, some dramatically. SPM goes down from above to just below average and is especially weak in Linkages and Interest (only 1·7 for MOLS students). Maths 2A maintains its impetus for MAPS students but loses the MOLS students altogether (Problem Sheets too difficult, Linkages low, Interest low). Maths 4A declines a little on most scores; and the Maths Majors gain interest while the Physics Majors lose interest.

Specific ratings of *lectures* changed in a similar fashion, though Chemistry improved on all five scales, especially on Interest, and SPM dropped even further. *Tutorial* ratings were all very high in the first term (except for Maths Majors on SPM), but only stayed high for Chemistry. Physics dropped a bit in Term 2, especially for the Chemistry Majors; and Maths 2A seemed to have lost the support of the Chemistry Majors altogether (1·0 for Interest). Nearly all the Physics and Chemistry *books* used by students received high ratings, only one 'set' book being deemed inappropriate. The Stereochemistry Programmed Text was also highly rated though most students complained of having too litle time to get through it.

It is difficult to disentangle two separate effects in looking at this data. First, we are looking at variations in the quality of teaching and the appropriateness of the courses offered; and secondly we are looking at variations in student motivation and interest according to how relevant for their intending major they perceived their course to be. It is possible that the two effects are interactive as it is easier to teach motivated students.*

2.2 Open-Ended Comments

Responses to a request for the best and worst features of the courses were quite varied, but certain items recurred frequently and are given below:

	Best	**Worst**
Term 1		
SPM	Lecture Presentation (12) Films and Demonstrations (5)	Laboratory Work (5)
Maths 2A	Lecture Presentation (7) Lecture handouts on vectors (5)	Lecture Presentation (12) Vectors and Mechanics (8)
Maths 4A	Analysis Lectures (6)	Linear Algebra Lectures (9)
Chemistry	Tutorials (4) Entropy (5)	Thermodynamics Lectures (7) Conciseness of Spectroscopy Course (6) Spectroscopy Lectures (5)
Term 2		
SPM	Tutorials (5)	Lecture Presentation (13)
Maths 2A	Lecture Presentation (6) (MAPS students only) Tutorials (4)	Lecture Presentation (9) Vectors (5)
Maths 4A	Analysis Lectures (5)	Linear Algebra Lectures (4)
Chemistry	Lecture Presentation (4)	

---

* Some of the interview evidence analysed since presenting this paper has been useful in revealing the relative significance of these two factors and the nature of their interaction.

The most notable characteristic of this list is the predominance of lectures under both 'best' and 'worst' features. This suggests that for most students the lectures are the most significant aspect of the course, a conclusion which is confirmed by the closeness with which trends in whole course ratings follow trends in lecture ratings (cf Section 2.1 above). A large number of students also took the opportunity of making comments in the open sections of the questionnaire which followed the rating scales on lectures, tutorials and courses. These were amplified and put in context during the course of the subsequent interviews; and are therefore reported along with the interview data in Section 4.

2.3 Topic Charts

The third questionnaire concluded with topic charts in which students were asked to comment on the novelty and difficulty of each topic (a topic usually being about a week's work); and to state which topics they spent most time on, and which they would have liked more help on. Features that appeared significant are listed below:

SPM – The only topics which significant numbers of students failed to grasp in Term 1 was the Schrodinger Equation (7 out of 25) and Wave Functions (8); though 4 non-chemists (subsequently identified as without 'A' level chemistry) failed to grasp the more chemical topics, Orbitals, Periodic Table and Bonding. These were also the most novel topics and those on which most help was requested (by over half the students in one case). Rather more topics caused difficulty in the second term, when 5 or more students failed to grasp 5 out of 9 topics. In both terms the allocation of study time seemed relatively even except for a marked fall in the final week or so.

Maths 2A – The difficulties arose at the end of the first term and in the second half of the second term, there being 4 topics out of 18 which about half the students failed to grasp. It also appears that students spent progressively less time as the course proceeded, the most studied topics coming at the beginning of each term.

Maths 4A – Most of the difficulties arose with the last 5 topics of the Linear Algebra, 4 of which half or more students failed to grasp.

Chemistry – Chemical Potential stands out as the one topic which several students failed to grasp. Many also requested more help with Spectroscopy and with other 'stereochemical topics'.

3 Student Allocation of Time as Revealed by Study Timetables

At three times during the Prelim Course students were asked to fill in study timetables to show how they had allocated the time they spent on academic work, both between courses and between different kinds of work. The academic activities were grouped under four headings, each of which was further subdivided as shown over:

Timetabled work – lectures, tutorials, skill sessions, problem classes, labs, quizzes.
Set work – problems, assigned reading, writing up lab reports.
Voluntary work – reading,* writing up lecture notes, writing own notes from books, exercises and problems, revision.
Other work – discussion with other students, scientific meetings, etc.

The overall trend is shown by the median scores for the four groups of students (in hours per week):

	Term 1 Week 6	Term 1 Week 9	Term 2 Week 5
Maths majors	44·0	34·0	32·5
Physics majors	30·6	27·5	30·0
Chemistry majors	33·3	33·0	30·1
Other majors in MOLS	35·0	35·5	31·8

The student estimated error is about 2 hours a week, and the last timetable was completed before revision had begun for the Prelim exams at the end of Term 2. In each week about half a dozen students did more than 42 hours and half a dozen did less than 27 hours work. Overall 4 students averaged more than 42 hours and 3 averaged less than 27 hours.

More detailed analysis showed that the enormous industry of *Maths Majors* in the first part of the course was largely due to the time spent on set work. (The median time spent on set work for the whole Maths group was 17·5 hours as compared with 8·0 hours for the Physics Majors.) The work was spread very evenly between the courses, and declined evenly as the overall total decreased. The *Physics Majors* worked the least but kept steady. The mean scores for three terms (33·4, 30·4 and 33·4) were higher than the medians because one student worked exceptionally hard (averaging nearly 49 hours a week) and two others averaged over 40 hours a week, whereas the three lowest averages (25·25, 27·6 and 28·0) were much closer to the median.

There were few differences between the two groups of *MOLS students* so we shall take them together. They spent more time on timetabled work than the MAPS students because of the heavy demands of Chemistry practicals; and as a result spent rather less time on set work and voluntary work. On the whole they worked evenly throughout the two terms, the reduction in Term 2 being entirely due to a drop in work on Maths 2A which is presumably related to the fall in interest already recorded. Even in the first term they spent less time on Maths 2A than the MAPS students and this was also true of SPM.

---

* Reading was subdivided into four categories: reading textbooks to gain understanding of course material; reading textbooks out of interest; reading ahead of the course; and reading lecture notes.

The following table shows the approximate percentage distribution averaged over the three study timetables for 5 different sub-groups (there are no sub-groups worthy of note amongst the MOLS students):

	% timetabled	% set	% voluntary	% other
MOLS students	60	21	14	5
Physics majors taking Principles and Perspectives of Science or Measurement Control and Computation (Electronics)	48	32	13	7
Physics Majors taking Maths 4A	45	27	22	6
Maths majors taking SPM	38	33	23	6
Maths majors taking Economics	34	52	10	4

It is worth considering whether this is the kind of distribution we imagine when we talk about a university education.

4 Student Perceptions of Prelims as Revealed by Interviews

The second interview, conducted at the end of the first term, explored students' initial reactions to being at university, their experiences of academic work in their first term and the influence of teaching on those experiences. There was an interview schedule which was adhered to loosely, and interviews lasted for between 20 and 30 minutes. Then the third interview, conducted at the beginning of the third term attempted to discuss motivation and the reasons for working or not working on various courses as well as continuing the exploration of issues of teaching and learning. There was a short list of topics but no interview schedule. Both interviews were taped (with the students' permission) in order to avoid note-taking.

The analysis of these interviews is still in progress, so I will have to restrict myself to reporting on two issues which have already emerged; the problem of feedback, and individual differences in student motivation.*

---

* Issues emerging from work subsequent to the presentation of this paper were: coherence, relevance and linkage; the content and pacing of courses; perceptions of 'good' and 'bad' lecturing; and 'good' and 'bad' tutoring. These will be included in the final report of the study.

4.1 The Problem of Feedback

The second interview devoted some time to the communication of aims, expectations and standards; and it emerged that, although with a few notable exceptions tutorials were strongly welcomed, many tutors failed to give students the kind of support for their independent work that they felt they needed. A third of the MAPS students and two-thirds of the MOLS students were not clear what was expected of them in terms of their private study and an even larger proportion were unclear about the quality of work expected. A half thought that feedback from lectures and tutors was inadequate. Some typical comments are given below:

'Probably not. They just tick it right or wrong. They don't normally comment on them at all. I think it would be helpful if they commented on our particular methods of doing these problems.'

'I suppose we were just expected to do as well as we can. . . . When it comes to tutorials we just hand in any work that we've done and get it back the next tutorial. They just say that it was all right.'

'Basically you're setting your own standards and that they pull you up if you're falling below what they think is a reasonable standard. If they think you are not doing enough problem sheets then they will tell you.'

'We're not told whether our work is good or bad – especially for the analysis part. It's just marked right or wrong and that's it. . . . We're not given any grading at all.'

Whereas some faculty may see marking in terms of maintaining student workrate or giving early warnings of possible failure, nearly all students saw it in terms of guiding their own self-evaluation. During the course of the first term they tried out a number of strategies: working over lecture notes, putting a lot of effort into set work, reading recommended books, and following up ideas of their own; and looked for cues from the tutor as to whether their approach to the course was on the right lines. But the tutor, more conscious of the problems of his own role, was more concerned with detecting student misunderstandings or non-understandings and remedying them. Once a deficiency was located, either by questioning or by failure to do a problem, the tutor's job was to cure it. This could be done interactively (the good tutor) or by lecturing (the bad tutor), but in either case the tutor's behaviour often suggested to the student a 'hidden curriculum which may well have been quite different from the tutor's own personal views. The student 'read the situation' and began to make up his own mind about what counted as important. So imperceptibly the declared intentions of lecturers and course documents gave way to a reality constructed from these often unplanned encounters. Whereas the support tutors gave for teaching, ie the lecturers and course books, was usually strong, the support they gave for learning, ie helping the student to create his own conceptual structures, to work independently and to know where he was going, was often very weak. It could be that in our attempts to avoid either railroading or cosseting the student we can easily give the impression that we don't really care.

4.2 *Individual Differences in Motivation*

Nearly all students wanted two things from their courses: an appropriate level of demand – they welcomed some challenge but nevertheless wished to be 'in control' of their work, rather than falling behind; and the arousal and maintenance of interest – they looked for and enjoyed certain types of academic experience but deplored others. Often these two requirements were closely linked, but the nature of the linkage varied greatly from student to student. Whereas for some students interest arose out of competence and they found it difficult to get interested in a topic before they felt they had more or less mastered it, for other students competence derived primarily from interest: they became highly competent in those topics which interested them and tended to ignore others. A third group were somewhere in between with a much more finely balanced interaction between competence and interest. For convenience we have titled these groups course-focussed, interest-focussed and intermediate; but in the characterisation which follows it should be remembered that these are 'ideal types' and the distribution should not be too rigidly conceived. Although it was the course-focussed students who were most concerned about feedback, that concern was by no means confined to that group. Nor were course-focussed students exempt from 'giving up' in certain situations where they had totallly lost interest.

Some typical quotations from interviews with interest-focussed students are given below:

'When I grew tired of certain aspects I didn't let it worry me. Certain aspects I pressed on with and did very well. . . . Essentially my work load is what I set myself. . . . Generally the subject matter that appeared in the first term did strike a chord in me and I felt inclined to follow it up. . . . Perhaps I shall approach this work again in the future but under my own steam and in a way which will maintain my interest.'

'My interest lies entirely outside of everyone else I meet and they are either not bothered about work at all or just want to learn the stuff in a set way.'

'If I were really interested I would try to read more about it and possibly go to see the tutor again depending how deeply interested I was. More so in the case of chemistry; sometimes in SPM and maths I just gave up. I just couldn't be bothered. . . I am also glad my SPM and maths tutors weren't so work conscious. . . . I wasn't interested enough to get the physics firmly implanted in my brain.'

The degree of independence shown by this group varied from those who saw the course as generating the interest to those who merely saw it as a framework within which they pursued their own personal interest. The former had high expectations of initiatives from tutors:

'I think the tutorial and the tutor is the most important part of it. If the tutor can engage your interest and keep it on the subject, then that's crucial.'

Whereas the latter tended to rely on the tutor more for moral support and discussion. For example, the same student who described an unusually perceptive tutor's response to his needs:

'After a short part of the Autumn term, I asked him if my work was satisfactory and he said "yes". . . . I suppose it was a sort of mutual respect. . . . He gave me the "sack" from attending tutorials, told me not to bother, and recommended some extra reading . . . told me to get on with it and go up and see him whenever I felt so inclined . . .'

also indicated a need for considerable tutorial support:

'When a professor says this to you, well you can't be doing too badly, and it encourages you to push yourself a bit further and not to be a bit anxious. . . . The use of "tutorials on demand" is good when you have problems but there is also a very strong place for a fixed tutorial, so to speak, so that a tutor can get things from you to test your progress.'

Course focussed students, on the other hand, tended to see the tutor as the servant rather than the master of the course. When asked if SPM tutorials enriched the course or engendered any interest one student replied:

'Not really but I have learned some things that I wouldn't have from the lectures. It helped me clarify a few points which is really what they're supposed to do.'

In general these students were concerned with competence and with fulfilling other people's expectations rather than their own:

'I stick to the syllabus and don't go beyond it.'

'I thought I'd spend a bit more time on maths; that was my weak point.'

'Not so much interested in it as I was good at it.'

'I thought the good part of the chemistry course was the quizzes afterwards.'

'My voluntary work was usually anything that wasn't made clear in the notes. Having to look in textbooks and things to make it a bit clearer.'

'There wasn't much time for following interests. It was just a case of keeping on top of the work that was being expected of us.'

'I originally decided that I was going to do every problem sheet, every problem on every problem sheet, but I didn't.'

'It's the same with everything. If I can do it, I'm interested in it.'

There is no time to describe any of the students in the intermediate group, so let me conclude by emphasising the significance of this typology for the way we design our courses and run our tutorials. Do we try and take into account the needs of both types of student?

Report on a Study of Staff Perceptions of Promotion Procedures and the Role of the University Teacher

Richard Startup

For a number of years I have been investigating the role of the university teacher. This has involved me in a detailed analysis of the various activities of staff within a particular provincial university, viz teaching, research, administration and additional professional activities. I have also been concerned to study how members of staff see their own role and also how undergraduates perceive the job of the university teacher. It will be readily appreciated that students and staff differ in the extent to which they have knowledge of particular staff activities (eg knowledge of research) and it can be shown too that there are systematic variations in the importance which they attach to the various activities of which they have knowledge (see Startup 'a').

In this paper attention will focus on a particular area of the wider investigation, namely staff perceptions of the procedures for promotion to the grades of senior lecturer and reader. This is a particularly important area of inquiry because it has to do with some of the incentives and rewards of academic activity. Inevitably rewards serve to channel activity and they can help or hinder the university in the achievement of particular stated objectives.

In the provincial university studied promotion is almost invariably internal. Decisions about individual promotions are taken by senior lectureship and readership committees. Only those who have themselves attained the relevant grade (or ones higher) are eligible to sit on the appropriate committee. The formal criteria link promotion to a readership with exceptional performance in research, while the senior lectureship is associated with achievements or responsibilities in connection with teaching, research, administration and other professional duties. These are the formal criteria, but here we are centrally concerned with how staff see the process and some of the resulting consequences of their perceptions.

Method

In connection with the study there have been three periods of data collection. Two surveys have been conducted. One was a survey of all the staff at the particular university. One hundred and ninety university teachers completed questionnaires, giving a 52% response rate. In addition, a survey of one in two second-year students was conducted and 321 (70%) responded. For both these surveys it can be shown that the samples are reasonably representative along certain important dimensions for which information is available from university records. (For example, when the distribution of the sample of staff between the two sexes, the four faculties and the four grades is compared with that in the population, in no case is the difference greater than that which results from random fluctuations.)

The final phase of data collection has consisted of interviews with staff. A department was selected from each of the four faculties and every member of staff was interviewed. In addition, four other heads of departments were selected and interviewed. It was felt that a minimum of approximately eight departmental heads was needed to give significant insight into the perspectives of those occupying this key position.

The Staff View of Promotion Procedures

In the survey of staff the respondents were first asked to indicate how satisfied they were with the procedures employed in connection with promotion. It can be seen from Table 1 that only 14% said that they were satisfied with the promotion procedures. Fully 45% expressed varying degrees of dissatisfaction, while over one third had mixed reactions. Particularly striking is the fact that even among those who have themselves successfully surmounted promotion hurdles (27 senior lecturers, 9 readers and 20 professors) there were more who were dissatisfied (18) than satisfied (14) with the prevailing system and a further 24 had mixed reactions to it.

Table 1 Satisfaction with procedures for promotion

	Satisfied or Very Satisfied	Mixed (Balanced) Reactions	Dissatisfied or Very Dissatisfied	Others	N
Professors	5	6	8	1	20
Readers	4	4	1	0	9
Senior Lecturers	5	13	9	0	27
Lecturers	12	42	68	8	130
All	26	65	86	9	186*
per cent	14·0	34·9	46·2	4·8	100

* *Of the 190 respondents, four did not indicate status and they have been omitted from Tables 1–3.*

Doubts about the adequacy of promotion criteria have been expressed by academics in previous studies. Halsey and Trow have described how 76% of their sample of British university teachers agreed strongly or with reservations that "promotion in academic life is too dependent on published work and too little on devotion to teaching" (Halsey and Trow 1971, p 349). In order to understand this issue further, in the present study the respondents were asked to consider the significance of the quantity and the quality of publications in connection with promotion. It can be seen from Tables 2 and 3 that almost one half of the staff felt that too much credit was given for the number of publications and a similar proportion felt that too little credit was given for their quality. In each case, in addition, approximately one quarter of the respondents said that they did not know enough about the procedures to judge whether sufficient credit or too much credit was given. Only a minority felt both conversant with the criteria used for promotion *and* satisfied with those criteria. There appears to be widespread suspicion that what counts in connection with publication is what is most immediately quantifiable. (For a more extended discussion of the staff survey data, see Gruneberg, Startup and Tapsfield 1974.)

Table 2 Weight given to the number of publications

	Too Much	Reasonable	Too Little	Don't Know	Others	N
Professors	4	11	0	3	2	20
Readers	3	6	0	0	0	9
Senior Lecturers	14	8	0	5	0	27
Lecturers	67	25	2	33	3	130
All	88	50	2	41	5	186*
per cent	47·3	26·9	1·1	22·0	2·7	100

Table 3 Weight given to the quality of publications

	Too Much	Reasonable	Too Little	Don't Know	Others	N
Professors	0	8	6	4	2	20
Readers	0	7	2	0	0	9
Senior Lecturers	0	6	14	7	0	27
Lecturers	3	15	68	42	2	130
All	3	36	90	53	4	186*
per cent	1·6	19·4	48·4	28·5	2·2	100

This view also emerged from the interviews with staff. Senior staff – and particularly those with personal experience of promotion committees – sometimes qualified their remarks about quantity by stating that the quality of publications was also considered (eg by account being taken of the character of journals, whether articles had been scrutinised by referees etc). There proved to be a substantial number of staff who in fact favoured the system which was supposedly operative, ie one which associated the award of a readership with achievements in research and the senior lectureship with all round performance. Yet many felt that in their university this important distinction was not being adequately maintained and research weighed very heavily (even decisively) in connection with both types of promotion.

For a significant number in the arts faculty there was the feeling that the present procedures discriminated against them. The emphasis on numbers of publications was seen as favouring the more prolific scientists. In discussing with staff their conception of research activity there emerged important differences between the ideas of scientists and those of arts men and women. The scientists were more likely to accept the notion that in some sense they were contributing to an 'advance in knowledge' while the arts men preferred a description such as 'enlarging understanding'. Some arts men just did not see their scholarly activities as being as likely as those of the scientists to speedily generate a string of publications. Support for this view can be provided from the Robbins Report where it is stated that: "In general, in the humanities periods of production come much later than in the sciences. . . . The young scientist needs time to press ahead with his research when his imagination is at its liveliest; his colleague in the arts may be better occupied in wide reading than in research on particular problems and he may well need more time free from teaching when his crop is ready for harvesting in middle life. A wide connotation must be given to the term 'research' if we are to apply it to all fields of study." (Committee on Higher Education 1963, p 184.) Senior arts men who were interviewed were particularly inclined to argue for separate faculty lists in connection with promotion. No doubt in the university at present some allowance is made for the nature of particular disciplines in connection with individual promotions, but some felt that this may work haphazardly and inconsistently.

The main recommendation of those interviewed was that teaching should in practice weigh more heavily in connection with promotion. A common remark was to the effect that the quality of teaching was only being assessed indirectly. In fact there is no systematic appraisal of student reactions to teaching. It was widely believed that teaching only came into the reckoning through the medium of the advocacy of the professorial head of department, who was interviewed by the promotion committee. Many respondents felt that heads of departments differed in their powers of advocacy and in their knowledge of the qualities of their colleagues as teachers. Indeed some of those interviewed were worried (for a variety of reasons) because they felt that their own head of department was insufficiently receptive to feedback from students. One senior member of staff familiar with the workings of the committees said that he had never heard a head of department admit that a promotion candidate was an indifferent teacher. The net result was that the committee succumbed to the weight of a large number of (what the members saw as) significant publications.

When interviewed, the eight heads of departments were asked to estimate the extent of their own knowledge about their colleagues qualities both as teachers and as researchers. Interestingly, the responses were highly variable. One head felt that he had great knowledge in both cases, but the others could locate significant gaps in their knowledge. Three of the eight described their own knowledge of teaching qualities as pretty small using descriptive phrases such as 'pretty minimal', 'no knowledge' (qualified later to 'negligible') and 'not that much knowledge'. On the other hand, there was one head of department who felt that he had a much more intimate knowledge of teaching than research qualities. The heads referred to a number of sources of information about teaching including feedback from students, listening to a man talking about his work and observations made in departmental societies and postgraduate seminars. Only once did a head of department refer to direct exposure to a fellow member of staff in a formal teaching session and it seems that there is a convention that members of staff do not normally attend one another's classes. It was stated earlier that promotion committees assess teaching qualities only indirectly. Judging from these self-assessments, it seems that the information that they are likely to be provided with is not only indirect but also highly variable in extent and quality.

Despite the widespread feeling among the staff interviewed that teaching should carry more weight than it in fact did, only a minority favoured a systematic appraisal of student reactions to or assessments of teaching. What might be thought to be a relatively 'moderate' scheme was put to many respondents for consideration. Under this scheme systematically collected student reactions would be presented to promotion committees for their further evaluation and interpretation (and hence the committee could choose in their wisdom to reject out of hand or to re-interpret substantially the information). Yet this 'moderate' scheme and system for using student assessment was rejected by a majority of those interviewed. There was the persistent feeling that students were not good judges of teaching. Some respondents stated that their own assessment of their teachers had altered over time and this assessment would be different at the time of interview as compared with their own view during the undergraduate period. There was the not unfamiliar point that lecturers might choose to 'play to the gallery'.

Another area touched upon in the interviews was the question of the mobility of lecturers between universities. Certainly a majority felt that lecturers were too immobile. A minority advocated making senior appointments open and this they felt would be beneficial to the standards both of teaching and research. Yet despite their consciousness of the problem of immobility many of their colleagues took a different line; they felt that those who had served the university well over an extended period of time had the first claim on promoted positions.

Some Student Reactions to Teaching

Information on staff perceptions of promotion procedures has so far been presented. It is now appropriate to consider the student response to the teaching which they receive, in order to tease out some of the effects of promotion procedures upon the operation of the university.

The student survey revealed that satisfaction with the presentation and content of lectures was fairly high. Some students, however, felt that there was not enough consultation with them concerning the content of courses. In addition, a quarter were dissatisfied with the amount of individual help they received from staff. (For more information, see Startup 'b'.) What it is important to focus attention on here is not so much the fairly high level of satisfaction with teaching but the problem of rectifying or ameliorating inadequate teaching. It turned out that a substantial number of students (including many who felt that standards were generally high) were of the opinion that not much was done, when lectures and courses were poor. Some comments made a direct connection between bad teaching and financial incentives:

"I would agree with the PIB that students should have some influence on their lecturers' salaries since there is very poor feedback to *bad* lecturers."

The root cause of the problem was seen by some as being in the appointments system:

"The selection of staff should depend as much on their ability to impart knowledge in an interesting way, as on qualifications."

Others are conscious of the need for lecturers to be taught to lecture:

"All lecturers should attend a compulsory course in 'how to lecture properly – clearly and concisely'. This might improve the abysmal standard of lecturing of *some* staff."

Earlier reference was made to the fact that staff did not generally experience one another's teaching in the lecture – or seminar room. Some writers have drawn attention to the functional importance of this. Merton, for instance, writes that "the norm which holds that what is said in the classroom of universities, is privileged, in the sense of being restricted to the (lecturer) and his students, has (the) function of maintaining a degree of autonomy for the teacher" (Merton 1957, p 375). However, this same writer also states that some measure of observability of performance (by other staff) is required "if the indispensable social requirement of accountability is to be met" (ibid, p 376). Heads of departments are known by students generally not to attend the lectures of departmental colleagues. Yet they are in the position of higher authority. Hence they can be seen to fail in the task of maintaining an adequate standard of lecturing.

"If the professor of each department were accessible to all students who wished to see him, he would get some idea of the proficiency of his lecturers and give them lecture courses suitable to their individual talents."

It can be seen that students felt that substantially more could be done to raise the standard of teaching to a uniformly high level. The activity (or non-activity) of the departmental head is a vital factor in this connection. Not only is the head the medium through which information regarding teaching qualities is transmitted to the promotion committee, but he is also seen by students (and others) as being responsible for the maintenance of teaching standards. Yet as we have seen, the knowledge which some heads possessed as regards teaching qualities was admitted to be minimal. In addition, teaching was not believed by most staff to be the decisive factor in connection with promotion. Hence the pressures from this source on the poor lecturer to improve his performance were not great.

Student and Staff Perspectives on Lecturing

When one talks about good or bad teaching or lecturing, certain features to which reference might be made are relatively straight forward, eg an inaudible lecture is a bad one, a lecture which is not meaningfully related to the previous knowledge or experience of the class is also a bad one. However, there are aspects of teaching which are more subtle and more apt to lead to differing judgements. The interviews with staff and the survey of students have helped to draw out one or two features of lecturing which can now be indicated (no doubt there are many more), where there is considerable scope for disagreement. It becomes clear that, if systematic assessments of teaching by students (or others) were to be instituted in connection with staff promotions, considerable thought would need to be given to the way in which evidence should be categorised and interpreted.

From the data collected, it emerged that some lecturers gave courses which were really summaries of smaller or larger amounts of published materials. Some students considered a mastery of the material given in the lectures, together with a minimal amount of additional reading, would be sufficient to do reasonably well in the examinations. They were not always disappointed in this expectation. Other lecturers gave courses which were intellectually challenging critiques of published materials, which would be extremely hard for students to absorb unless they did a considerable amount of reading in support of the course. If they attempted to tackle a course of the latter type, with the same approach that would be appropriate and adequate (as they saw it) for the former type, they might fail to do themselves, or the course, justice. There is clear evidence of confusion in students' minds regarding what lectures are *for*, and what lecturers are attempting to *do* in their courses. It can be readily seen that these differences in approach to lecturing pose problems for the interpretation of student reactions. It must not be thought that this example concerns merely the content of lectures and not their presentation. If a lecturer is presenting an intellectually challenging critique his manner and his relationship with the material (and the feelings he evinces) may be expected to differ from the case when he is giving a summary of material.

A second problem area can be highlighted. This time it concerns systematic differences between the perspectives of many lecturers and those of a substantial body of students. It proved to be the case that many university staff intended in their lectures, to do justice to the most central theoretical notions within their disciplines. They valued the mastery of theoretically important ideas for itself and they saw themselves inviting students to join them in what they saw as intrinsically rewarding intellectual activity. Evidence has been presented elsewhere (Startup 'c') that the students' reasons for coming to university, involved an emphasis on the subsequent occupation which will be pursued. They were sensitive to the practical and occupational significance of what they were taught. Hence the objection that 'lectures are too theoretical and divorced from real life'. Only a minority placed the greatest emphasis on mastery of an academic subject for itself. Thus, we find a serious discrepancy between what many students expect and require of the university teacher and what he intends to provide.

These kinds of difference in the values of students and staff pose problems for the systematic use of student assessment. It will be readily appreciated that one needs a clear statement of what lectures are for and what lecturers are trying to do, before one can determine how well they perform their task. What emerges clearly from this study is that there are significant aspects of the art of lecturing about which there is no clear consensus embracing staff and students.

Concluding Remarks

This paper has been centrally concerned with the staff view of promotion procedures; it has not been our purpose to present an 'objective' view of the mechanics of those procedures. Perhaps beliefs widely held by staff are false. Indeed, one of the departmental heads interviewed stated that the workings of the Senior Lectureship Committee came as a surprise to many who joined it. Attention has been given here to staff perspectives on the grounds that widely held beliefs shape actions and generate effects whether those beliefs correspond to 'reality' or not. Given that false beliefs frequently have undesirable effects, those responsible for promotions may be encouraged to set about actively combating any misconceptions. It seems likely that this will require the promotion committees to operate more openly and with more detailed clarification of criteria. It may also require the participation of all staff rather than just the minority consisting of those already promoted.

Some arts men felt that the emphasis on publications discriminated against them. Given the differences in the type of scholarly activity in which arts men are engaged from that pursued by scientists, it would seem sensible to institute some formal separation of promotion procedure for different subject groups. It may well be that the present promotion procedures do this effectively on a case by case basis. If this is so, making the operation more open and formal may help staff to become aware that their particular kind of scholarship is appropriately recognised.

A central difficulty with the present system for promotion is the indirect assessment of teaching qualities. The departmental heads gave every evidence of varying substantially both in their knowledge of their colleagues teaching qualities and their powers of advocacy before a committee. Neither was it felt to be likely that they would admit that a candidate was a bad teacher. This may be an unfortunate consequence of the tendency for their role to be defined as that of advocate and sponsor. If teaching is assessed in this way, it would hardly be surprising if the altogether more tangible books and off-prints carried the day.

Despite the felt need to weight teaching more heavily in connection with promotion, only a minority of staff who were interviewed favoured the use of a systematic appraisal of student reactions to or assessments of teaching. Many doubts were expressed about undergraduate ability to recognise good teaching and it is true that judgements made by an individual may change over time. Nevertheless, some of the arguments against the systematic determination of student reactions look strained. It seems likely that considerable staff unease flows from the fact thta there would be an accretion of power to students if their views carried weight. Some evidence has been provided here that there are differences of values between students and staff concerning what constitutes good and bad teaching (no doubt there are important areas of agreement as well). There is no avoiding the point that where there are differences of values, if students gain more power, their view regarding appropriate teaching is that much more likely to prevail. Yet students are effectively the only 'consumers' of teaching and their reactions *must* be relevant to an assessment of teaching qualities.

References

Committee on Higher Education (Robbins), *Higher Education*, London, HMSO, 1963

M M Gruneberg, R Startup and P Tapsfield, "A Study of University Teachers' Satisfaction with Promotion Procedures", *The Vocational Aspect of Education*, XXVI, No 64 (Summer), 1974

A H Halsey and M A Trow, *The British Academics*, London, Faber and Faber, 1971

R K Merton, *Social Theory and Social Structure*, New York, The Free Press, 1957

R Startup, 'a', "How Students See the Role of University Lecturer", *Sociology*, 6, 2 (May), 1972

R Startup, 'b', "Student Satisfaction with Academic Services", *Educational Research*, 14, 2, 1972

R Startup, 'c', "Why Go to the University?", *Universities Quarterly* (Summer), 1972

Student Evaluation of Teaching at North East London Polytechnic

Philip S Bradbury & Paul Ramsden

We describe in this paper an attempt which began this year at North East London Polytechnic to put into effect means by which lecturers can obtain systematic feedback on their teaching, using questionnaires. As far as we know, no other British institution of higher education has attempted a scheme such as this on a large scale, although the practice is common elsewhere, especially in America (Flood Page 1974).

The questionnaire currently in use is one outcome of a feasibility study carried out during 1973/74 on questionnaire methods of course and teaching evaluation by students. It was designed after interviews with staff and students on a number of courses at NELP, and a preliminary version was tested by eleven staff teaching on seven courses broadly representative of the polytechnic's range of full-time, part-time, degree, sub-degree, vocational and non-vocational courses in different faculties (Ramsden 1975). Checks on the reliability of this questionnaire and on its applicability to different courses proved satisfactory. Most of the users said that they had found the exercise helpful and would be making changes to their teaching as a result.

Current Use of Teaching Evaluation at NELP

The present questionnaire comprises 23 statements about teaching, against each of which students rate their lecturer (see Table 1). In a second section of the questionnaire students are asked to rate the importance of each of the 23 statements, thus indicating their definitions of desirable teaching attributes and qualities (see Table 2). It is intended that the second section should be used to interpret the results of the first.

In the Autumn Term 1974, we offered a service to all NELP teaching staff: the supply of questionnaires and, optionally, the confidential analysis of the results. Staff were circulated with a leaflet describing the questionnaire and explaining why we thought they might want to use it. By Easter 1975, 94 requests had been made by teachers and 3,000 forms had been sent out. A wide range of courses and departments were represented, and an equal proportion of senior to junior staff asked for questionnaires. About half of the staff asking for forms sent them back for analysis; the remainder analysed the results themselves. The analysis provided by the project consisted of a simple description, in graphical form, of the number and percentage of replies to each question. Any further interpretations were left to the teacher himself.

Table 1 Student Feedback Project – Section I: Feedback Questionnaire

Please indicate, by circling the appropriate number in each case, your agreement or disagreement with each of the following statements:

The Lecturer	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1 Is clear and understandable in his or her explanations	1	2	3	4	5
2 Stimulates students to think independently	1	2	3	4	5
3 Makes a genuine effort to get students involved in discussion	1	2	3	4	5
4 Presents material in a well-organised way	1	2	3	4	5
5 Makes good use of handouts, eg duplicated lecture notes, examples of problems, reading lists	1	2	3	4	5
6 Is sensitive to the feelings and problems of individual students	1	2	3	4	5
7 Is enthusiastic about his or her subject	1	2	3	4	5
8 Stresses important material	1	2	3	4	5
9 Makes constructive and helpful comments on written work	1	2	3	4	5
10 Adjusts his or her pace to the needs of the class	1	2	3	4	5
11 Gives a good factual coverage of the subject matter	1	2	3	4	5
12 Shows a good sense of humour	1	2	3	4	5
13 Writes legibly on the blackboard	1	2	3	4	5
14 Returns written work promptly	1	2	3	4	5
15 Shows a thorough knowledge of his or her subject	1	2	3	4	5
16 Shows the relevance of his or her subject to the work you expect to do when you qualify	1	2	3	4	5
17 Points out the links between his or her subject and related subjects	1	2	3	4	5
18 Is readily accessible to students outside formal classes	1	2	3	4	5
19 Encourages students to express their own opinions	1	2	3	4	5
20 Is always well prepared for his or her classes	1	2	3	4	5
21 Is punctual and reliable in his or her attendance	1	2	3	4	5
22 Tries to link lecture material to laboratory work/practical work/fieldwork/seminars	1	2	3	4	5
23 Can be clearly heard	1	2	3	4	5

Please make sure you have answered all the questions

Table 2 Student Feedback Project – Section II: Importance Ratings

Please indicate how important *you think it is that lecturers teaching on this course should show each of the following characteristics. Do this by circling the appropriate number on the five-point scale [ie* 1 *if you think the characteristic is of* little *importance and so on up to* 5 *if you think it is of* great *importance*].

Lecturers on this course	Of Little Importance				Of Great Importance
1 Should be clear and understandable in their explanations	1	2	3	4	5
2 Should stimulate students to think independently	1	2	3	4	5
3 Should make a genuine effort to get students involved in discussion	1	2	3	4	5
4 Should present material in a well-organized way	1	2	3	4	5
5 Should make good use of handouts eg duplicated lecture notes, examples of problems, reading lists	1	2	3	4	5
6 Should be sensitive to the feelings and problems of individual students	1	2	3	4	5
7 Should be enthusiastic about their subject	1	2	3	4	5
8 Should stress important material	1	2	3	4	5
9 Should make constructive and helpful comments on written work	1	2	3	4	5
10 Should adjust their pace to the needs of the class	1	2	3	4	5
11 Should give a good factual coverage of the subject matter	1	2	3	4	5
12 Should show a good sense of humour	1	2	3	4	5
13 Should write legibly on the blackboard	1	2	3	4	5
14 Should return written work promptly	1	2	3	4	5
15 Should show a thorough knowledge of their subject	1	2	3	4	5
16 Should show the relevance of their subject to the work you expect to do when you qualify	1	2	3	4	5
17 Should point out the links between their subject and related subjects	1	2	3	4	5

18 Should be readily accessible to students outside formal classes	1	2	3	4	5
19 Should encourage students to express their own opinions	1	2	3	4	5
20 Should always be well prepared for their classes	1	2	3	4	5
21 Should be punctual and reliable in attendance	1	2	3	4	5
22 Should try to link lecture material to laboratory work/practical work/fieldwork/seminars	1	2	3	4	5
23 Should be clearly audible	1	2	3	4	5

Please make sure you have answered all the questions

The Reliability and Validity of the Results

Data from the staff who have used the questionnaires have been available for more detailed analysis. Our analysis suggests that results from the questionnaires are sufficiently accurate and consistent to be of use to the individual lecturer in improving his teaching.

First, the results seem generally to be accurate reflections of how students perceive their lecturers' teaching. One of the lecturers, for example, gave the questionnaires to three groups of students, each group following a different course (two degree courses and a one year professional course) but to whom he taught the same subject (mathematics). Although these groups varied in the degree of 'generosity' of the ratings given (the professional course students tending to be most generous), the different groups picked out the same 'strengths' and 'weaknesses'. Subsequent discussion with the lecturer revealed that he had been able to identify these, and take appropriate action where he thought necessary.

Secondly, the results clearly distinguish between different teachers, and different profiles of perceived strengths and weaknesses emerge for different lecturers. A rank correlation of scores for the first part of the questionnaire for two different lecturers, both teaching social psychology, yielded a coefficient of 0·01. In another typical case, the correlation coefficient between scores from the same group of students for two different lecturers was −0·23.

Thirdly, the results appear to be highly reliable. Five lecturers agreed to undertake a reliability retest with their students between seven to fourteen days after the original evaluation. In two cases the conditions were controlled such that *exactly* the same students were involved in both test and retest. In the remaining three cases there were slight differences due to absenteeism. (It is clearly difficult to control completely for absenteeism and at the same time preserve the anonymity of the students, since the students would have to make some identifying mark on their questionnaires.) In no case did the reliability coefficient fall below 0·79 for section I of the questionnaire or

Table 3 Using the Feedback Questionnaire to Change Teaching Strategies – an example

Question	Students' relative rating of lecturer	Students' relative rating of importance	Lecturer's relative rating of importance	Probable source of conflict	Action
Presentation (Q4)	low	high	high	lecturer's style	Spell out structure of material more carefully
Written work (Qs 9 & 14)	low	low	high	none handed in	Agree more clearly laid down guidelines for undertaking assignments
Accessibility (Q18)	low	low	high	lecturer based at another building	Lecturer spends lunchtimes with students
Handouts (Q5)	low	low	high	not used very much	Increase use where appropriate
Factual coverage (Q11)	low	high	low	not a course objective	Regularly restate course objectives

below 0·77 for section II. The data, however, suggest equally strongly that the questionnaires cannot properly be used for comparing teachers, either against some 'standard' or against each other. It is clear, for example, that student expectations vary between student groups. Data from section II of the questionnaire reveal that there are differences in the ways that students from different faculties expect to be taught (Ramsden 1975), and also that students in the first year of a course tend to be more 'optimistic' in their expectations than do students in subsequent years. Also, as we have seen above in the case of the mathematics lecturer, different groups of students vary in the degree of 'generosity' which they show to the *same* teacher in their ratings. Finally, there are many sources of correlated bias which affect the results, including different teaching methods and the time of administration of the questionnaire.

A Case Study of the Use of the Questionnaire to Improve Teaching

One of the authors has used data from the questionnaires to change and (he hopes) improve his own teaching strategies using results from his own students (Bradbury 1975). By *ranking* the statements (from that which most students agreed with as being shown in his teaching to that which fewest agreed with), what students perceived as relatively weak points became apparent. Similarly, results from the second section of the questionnaire were used to establish how important students felt the perceived weaknesses to be. Taking into account his own teaching priorities the author was able then to use the data to make a number of specific changes to his teaching. The details are outlined in Table 3.

On Confidentiality and Voluntary Use

We feel it is most important that the use of student evaluation questionnaires should be entirely voluntary. The main purpose of the exercise is, after all, to help an individual to find out more about how his students see his teaching; he is unlikely to get much from the results if he feels that he has been coerced into using the questionnaires. But the need to persuade as many lecturers as possible to use the questionnaires is a real one, for good teachers, who may be most confident about their teaching, are probably the ones who make most use of them at the moment. Pressure from students to use the questionnaires might be more acceptable than institutional coercion, although it still raises difficulties. The situation has arisen, for example, in which students on a particular course have demanded that all staff should use the questionnaires, following their use by one teacher. In this case, students had felt that the 'good lecturer' had used the questionnaires, while the 'bad lecturers' ought to use them.

We cannot accept that the results of evaluation questionnaires should be other than strictly confidential to the lecturer concerned; although, as we suggest below, it might be acceptable to use the data from a large number of forms for limited institutional purposes, so long as individuals were not identifiable. In the last analysis, the individual teacher is the best judge of what his own results mean; and many lecturers who would perhaps be prepared to use this method of evaluating their teaching would not do so if they felt that there was any possibility of their own results finding their way into the hands of heads of departments or other senior staff.

Monitoring Procedures

What are the effects on students and staff of using feedback questionnaires? A programme of interviews and discussions with teachers and students has recently begun to assess these effects. It is important to know, for example, whether there is any possibility of damaging the relationship between students and staff by using these questionnaires.

On the limited amount of information we have so far gathered, it appears that students are generally pleased to complete the questionnaires. This may, however, be an effect attributable to the novelty of the exercise. An important worry which some students seem to have is that their individual comments may not be anonymous. Every care, then, must be taken to preserve students' anonymity – we recommend, for example, that lecturers are not present in the classroom when the forms are being completed – if students are not to think that their responses might be used against them. This is one of the reasons for discarding open-ended questions on our current questionnaire. A feeling that the responses may not be anonymous is also likely to affect the validity of the results. Another problem, so far as students are concerned, is that they may think that the exercise has been worthless if they do not perceive constructive changes in areas of teaching weaknesses. At present, however, we do not have enough information to decide whether this is happening or not.

We are also studying the reactions of teachers to the questionnaires. Several staff have been interviewed and a letter will be sent to all users to discover their opinions. It appears important that lecturers should feel able, having used the questionnaires, to make or plan specific changes to teaching in the light of the results. The questionnaires also ought to tell staff something about their students' views which they did not already know, rather than merely confirm suspected strengths and weaknesses – although this may be an added benefit. Nearly all of the teachers so far interviewed have agreed that the results have encouraged them to make changes, and most have found out something they did not know about students' opinions of their teaching. Important side effects for some lecturers were that using the questionnaires had concentrated their attention on teaching aims, and that, because asking for students' assessments is an expression of a teacher's confidence in his students' judgement, staff–student relations had been improved.

Institutional Use of Feedback

The evidence already presented above clearly suggests that the questionnaires cannot be used for comparing teaching abilities. This being the case it would be quite inappropriate for results to be used for decisions about staff tenure and promotion.

There are, however, arguments for an association between staff development programmes and feedback procedures. For example, what is a lecturer to do if he finds students consistently find him 'lacking' in aspects which both he and his students regard as important and despite all efforts to improve? This could happen even with relatively simple things like audibility and use of visual aids, and certainly with the more difficult teaching skills such as promoting group discussion. The institution clearly has a responsibility to provide facilities for those who wish to use them. Whether staff development programmes should be planned in the light of collective feedback data received by a central unit or through specific demands of staff who feel they need a particular facility is open to question. The former possibility could be regarded as a breach of confidentiality.

Conclusions

Student feedback questionnaires have been found most useful, from our own experience, as a means of self-evaluation by the individual lecturer. The use of the questionnaires should be voluntary and the results kept private, or the process may be a self-defeating one. It is doubtful whether the questionnaires can measure teaching ability, but they provide an excellent indicator of how students see one's teaching strengths and weaknesses; this information can be used to make changes to teaching. An important problem remaining concerns the ways in which those teachers who perhaps most need to use the questionnaire, and act on its findings, can be encouraged to do so.

References

P S Bradbury, "A question of feedback", *Aim* 1, 2–7 North East London Polytechnic, 1975

C Flood Page, *"Student Evaluation of Teaching: The American Experience"*, Society for Research into Higher Education, 1974

P Ramsden, "Polytechnic students' expectations of their teachers and the use of a student feedback questionnaire: a preliminary report", *Higher Education Bulletin* (forthcoming), 1975

Evaluation Studies of Six Courses in Geography, Electrical Engineering and Physics

Professor P J Black & Dr R J D Rutherford

This paper describes studies which were initiated by a group of lecturers who decided to develop work together on the evaluation of courses within three departments in the Faculty of Science and Engineering at Birmingham. Each evaluation study began with a completely voluntary contract between lecturer and evaluator; the latter from a different department, but in a related discipline, and therefore with some understanding of the subject matter but hopefully without biases or preconceptions as to how the material should be organised and presented. For four of the courses, both lecturer and evaluator were members of the group who were thereby working together to observe and analyse one another's courses. For two others, the lecturers concerned did not belong to the group. In the course of the work the group came to adopt, for the preliminary evaluation of a course, a general procedure which has five main stages.

The first stage was for the evaluator to observe a selected sequence or 'episode' of teaching, relating to a particular topic. This involved the evaluator 'sitting in' on a number (typically three or four) of lectures, and associated tutorial classes, practical classes etc as appropriate, and making some fairly detailed notes of his impressions. In the second stage, a group of six to eight students, selected on a random basis, discussed the course in the presence of the evaluator. The discussion was recorded for analysis later. The evaluator did not act as a chairman but tried to let the discussion develop freely between the students: he intervened only to introduce new topics for discussion in order to ensure that all relevant aspects of the course had been discussed.

The third stage was for the lecturer and the evaluator to design a course evaluation questionnaire which would be completed by the students and also by the lecturer himself. The views of the lecturer, the evaluator, and most of all the students were taken into account. The group tried to make the student discussion and the questionnaire as comprehensive as possible, and these covered the following issues: students' general attitude to the subject, their image of the lecturer, appreciation of

---

Note: This paper was presented at the Conference by Professor Black and Dr Rutherford. However, the full authorship of the paper includes Professor P J Black, Dr D E Evans, Dr W A Kimball, Dr R J D Rutherford and Dr J W R Whitehand.

the aims of the course, the transactions in the lectures, the use of textbooks, the perception of examinations, the work in tutorial classes, the perceived links with other courses and the previous knowledge assumed by the lecturer. Although the overall agenda was the same, the issues that came up differed from one course to another so that each questionnaire contained items which were specific to the course concerned.

In the fourth stage the students completed the questionnaire, generally at the end of a lecture period, and the results were analysed. In the final stage the lecturer and the evaluator would write a short report which was then discussed at one of the regular meetings of the group of staff.

In addition to this general procedure it was sometimes possible to discuss the results of a particular questionnaire with students and to video-tape a complete lecture. A number of courses have been investigated along these lines and each evaluation has tended to highlight quite different problems. These can be illustrated by describing a number of case studies.

Case 1 First Year Electrical Engineering Course A

This course was concerned with electrical circuit theory: in order to achieve the main aim, which was to help students to develop skills in problem solving, the lecturer spent almost all of the lecture time in dialogue with the class, posing problems and then proceeding to work through them by posing questions and responding to answers and suggestions from students. The discussions with students and the questionnaire showed that students claimed to have more than a purely vocational interest in their course and that they respected the enthusiasm and professional expertise of the lecturer.

Of the 43 questions which made up a questionnaire to students, 9 are shown, with their responses, in Table 1. The first two of these showed a difficulty with the recommended course textbook: this came as a surprise to the lecturer and is an important feature because of the lack of formal lectures. Several questions made it clear that the two hour problem classes associated with the lectures were relevant and potentially helpful, but questions 3 and 4 confirmed that particular difficulties mentioned by students in a small discussion group were shared by most of the class. Questions 5 and 6 arose out of a concern of the evaluators who observed that few notes were taken and who saw and heard evidence that the interactive lecture situation might be confusing to some students who might not follow the thread in the diverse arguments. It would be tempting to say that the response to question 7 provided further evidence of confusion, but this must be offset by question 8 which states the lecturers' view of the main aim of the course and shows that students had perceived it. Question 9 shows a possible difficulty about examinations, although it could be argued that this information was contained in the set problems rather than in those tackled in the lecture. These questions have been selected to emphasise the main problems arising from the evaluation. The overall impression of the evaluators was that the students wanted clearer and firmer guidance and support material to complement the involvement and stimulus of the lecture dialogues.

Table 1 Responses to Selected Questions for First Year Electrical Engineering – Course A

The numbers quoted are actual numbers of student votes under the four headings: strongly disagree, disagree, agree, strongly agree.

	s/d	d	a	s/a
1 Reading textbooks is essential to the full understanding of the electrical engineering courses	0	5	24	15
2 I have made little or no use of the textbook by (name of recommended textbook given)	6	12	20	6
3 When we don't cover all of the problems in the two hours nothing can be done about the ones that aren't covered	0	2	26	14
4 The lecturers explain many methods for solving problems but we are not taught how to choose the best method when we meet a new problem	2	12	27	2
5 I take very few notes during his lectures	1	16	19	8
6 I came out of most of his lectures feeling clear about the point of the lecture or discussion	8	24	12	0
7 He makes clear during the course what the course is supposed to achieve	0	28	15	1
8 He is trying to get students to use an electrical engineer's approach to tackling problems in the subject	0	4	29	11
9 The lectures make it easy for me to see the kind of problem I would be expected to tackle in the examination	3	29	11	1

Case 2 First Year Electrical Engineering Course B

This course was given to the same group of students as course A, and the two courses formed part of an integrated scheme of teaching which meant that they were to some extent complementary to one another and that they were linked to a common tutorial (problem class) system. The two courses overlapped in time but course B went on later so that the evaluation was completed several weeks after the one described above, and after a christmas vacation instead of before it. In this course the lectures were delivered in traditional style, and full duplicated notes were handed out beforehand. The student questionnaire again showed that they respected the enthusiasm and professional expertise of the lecturer but showed also (questions 1 and 2, Table 2) that the students felt they had a much better appreciation of the aims for this course.

The student discussion group seemed very satisfied with the lectures and the only criticism emerged in a suggestion, agreed by the group but rejected by those answering the questionnaire (question 3, Table 2). The evaluators tried to check on the possible effects of providing lecture notes beforehand but found no support for the usual fears about the effect of such notes (questions 4 and 5, Table 2). In discussion with the student group the evaluators inserted questions about the contrast of the lectures with those of course A, and students did agree that they might be helped by some encouragement to ask questions and (question 6) by being more actively involved during the lectures. The problem class system associated with the lectures in course A was also used for the lectures in this course: as a consequence of some student criticisms, some of which were revealed in the first study above, the method of running these had been altered and question 7 showed that the changes were seen to be helpful whilst question 8 confirmed that a positive feature which was appreciated for course A was similarly rated for course B. It was expected that question 9 would have received a more positive response than the corresponding item (question 9) in Table 1 because of the more explicit guidance provided and the better communication about aims. There is a significant change but the responses still show that there is a problem; this may arise because students did not relate the problems and other materials provided to examination policy. Questions 10, 11 and 12 explored the issue introduced in questions 1 and 2 of Table 2.

Although the two courses were very different in style, in neither were many students using the course textbook(s) although the lecturers expected and hoped that they would do so. It is not clear why question 10 (Table 2) gives such different results from question 1 in Table 1. It may reflect a shift in attitudes during the course, but it is more likely that it arose from the question being judged in the context of the course being evaluated.

The general effect of this course was in sharp contrast to that of the previous one – the aim, of understanding one specific area of electrical engineering, was quite different from the 'problem solving' emphasis of the first course and the method of presentation was clearer and more supportive. Students were satisfied mainly because the course was well taught, but there was a hint (eg question 6, Table 2) that it might have challenged and involved them more. It was evident to one evaluator who sat through successive lectures, one on course A and one on course B, that at the end of a lecture on course A students relaxed and the atmosphere amongst them suddenly altered to a quieter and more passive one. It is tempting to propose that each course suggests ways in which the other could be improved but the evaluation cannot specify solutions to the problems that it exposes and ought not to be used in this present instance to suggest that either course needs radical change in the direction of the other. The evaluation does serve to show typical clusters of interrelated problems which in several instances are unique to the detail of the presentation of the course concerned.

Table 2 Responses to Selected Questions for First Year Electrical Engineering Course B

		s/d	d	a	s/a
1	He makes clear during the course what the course is supposed to achieve	0	4	27	5
2	He actually achieves the main aims of the course	0	4	26	4
3	The lecturer should have gone more quickly in the first few lectures so that he could have spent more time on the later material	2	24	6	2
4	I usually make some notes of my own during lectures	7	10	19	0
5	One could follow this electrical machines work quite well using the notes only without attending the lectures	9	22	4	0
6	The lecturer should ask us questions to make us think and take part during the lectures	0	10	20	5
7	The system of handing in problems for marking is more useful to me than last term's system	1	5	25	4
8	The problems set are closely related to the lecture courses	0	1	31	4
9	The lectures make it easy for me to see the kind of problem I would be expected to tackle in the examination	1	16	18	1
10	Reading textbooks is essential to the full understanding of the Electrical Engineering Courses	4	22	10	0
11	I have made little or no use of the textbook by Hammond	2	5	15	12
12	I have made little or no use of the textbook by Stephens	2	6	16	11

Case 3 First Year Atomic Physics Course

In this course the lecturer used an interactive style in that much of the lecture time was devoted to discussion arising out of student answers to questions posed to the class. However, he used a wider range of question techniques than in course A above, using multiple choice questions for vote by the whole class and intervals for students to discuss a question in addition to questions to be answered by individual volunteers. There was also more support material: students received brief lecture notes after each lecture, a syllabus with detailed references to passages in a course textbook and a detailed list of aims. Some of the course work was left to private reading and the course was accompanied by problem classes based on problems some of which, being based on passages in the textbook, obliged students to consult it. Responses to questions 1 and 2, Table 3 shows that these lectures, although comparable to course A in style, achieved clarity with regard to aims comparable to that of course B. Question 3 showed that this could not be explained directly by student use of the list of aims issued, a point which is reinforced by the positive response to question 4: the lists referred to here were handed out to students attached to the list referred to in question 3. The question about textbooks matched that for courses A and B. No question was asked about use of textbooks because it was known, from the work handed in by students and from a previous questionnaire, that the course textbook was used by all students. From his attendance at some lectures, the evaluator was concerned that a minority might be taking an active part with the majority playing a purely passive role, feeling frustrated at being left behind and gaining little from the lecturer's attempts to encourage active thought in the lecture theatre. Question 6 showed a division about the need to take notes (as for course B above), perhaps related to satisfaction with the notes provided (question 7). The other four questions (8 to 11) showed that although the evaluators evidence (eg question 9) was accurate, the worry about the effects of this on the majority was not shared by the students (questions 10 and 11), and they did not lose track in the rather erratic progress which can arise in an interactive system.

Table 3 Responses to Selected Questions for a First Year Atomic Physics Course

	s/d	d	a	s/a
1 He makes clear during the course what the course is supposed to achieve	0	12	29	3
2 He actually achieves the main aims of the course that are explained for 1	0	7	34	1
3 During the course of lectures I have referred to the list of course aims of the Atomic Physics Course issued by the lecturer	4	26	13	0
4 During the course of lectures I have referred to the lists of reading and lecture material issued by the lecturer	1	12	29	3
5 Reading is essential to a full understanding of the atomic physics lecture course material	0	3	28	14
6 Taking notes during these lectures is essential to the process of learning and understanding	4	21	19	2
7 The lecture summaries provide an adequate record of the lectures	0	9	30	7
8 I come out of many atomic physics lectures feeling confused and unsure about the point of the lecture or discussion	5	28	11	1
9 Answering questions put to the class as a whole tends to be dominated by the most able in the class	3	10	14	18
10 Ideas are covered too quickly because the lecturer is talking with the most able students	5	28	8	5
11 His approach has helped and stimulated me to think creatively about physics	2	9	30	5

Views about assessment were not explored in this evaluation. Several of the questions in the questionnaire reflected criticisms expressed in the student discussion group which had tended to reinforce some of the doubts of the evaluator about the success of the course. However, the questionnaire responses seemed to be at variance with the views of the discussion group in that these criticisms received little support.

Case 4 First Year Course in Geography

This study concerned the evaluation of a first year course in geography, one of the eight foundation courses in geography at Birmingham. This was a course of some 24 lectures, given to about one hundred students. There were no tutorials specifically related to this course nor was there any set work to be handed in for marking. A reading list was issued at the beginning of the course. The procedure for evaluating this course was somewhat different to that outlined for the first three studies.

Before the course began the students were asked three questions relating to their reasons for coming to university to study geography. The students were asked to answer each question on a four-point scale, where four was always the most favourable response. The three questions were:

1 I am more interested in obtaining a degree than in the academic subject;
2 I see geography as part of a training for a specific job/career;
3 I am interested in geography as an academic study in itself.

The averages of the students' response to each of the three questions were 1·7, 2·6 and 3·2 respectively. (On similar questions in the other evaluations first year electrical engineers gave average scores of 2·0, 2·8 and 3·0 and first year physicists gave 2·0, 2·0 and 3·1 respectively.) The immediate conclusion is that these are very keen students and only six of the ninety-seven replies rated question 1 above the others. When these results were discussed with a group of students it seemed that 'the idea of coming to university' was important to them, and also the feelings of surprise at finding out how easy it was to fit in and belong to the university community.

The second questionnaire that the students were invited to complete was more specifically concerned with the effectiveness of the teaching and it included questions such as:

1 organises lectures well;
2 arouses students' interest.

The results of this questionnaire were discussed with a second group of students and it quickly became apparent that there was a wide divergence of opinion within the group as to what constituted good lecturing.

It was possible to video-tape the whole of one of the lectures and on viewing this the lecturer became convinced of the need to include a variety of activities during the lecture and particularly to provide a break halfway through. He was also concerned that the lecture appeared too smooth and polished with the result that the material seemed to be presented in a very cut and dried fashion. This experience suggested that the aims of the course, and in particular how those should be examined, should be considered. The lecturer had emphasised that in the first year examination simple recall of lecture material would play a small part, more important were the abilities to integrate different parts of the course, to discriminate between the relevant and the irrelevant when answering questions, and to show understanding and initiative in their writing. In past years, many students (15%) had failed to appreciate these aims and had failed the examination at the first attempt. This was a much higher failure rate than in the other seven foundation courses. At this stage the evaluator had attended about eight lectures and was feeling confident about his knowledge of the course and what was expected in the examination. He wrote out ten possible examination questions, but was informed by the lecturer that none of these were likely to appear on the final paper! A questionnaire with ten possible questions, five written by the evaluator and five by the lecturer, was drawn up and the students were asked to rate the likelihood of each of these questions appearing on the final examination paper, using a four-point scale. In fact the students failed to recognise whether a question was a possible examination question or not, and there was a total mis-match between students' perceptions and the lecturer's own judgement in six of the ten questions. There was a discussion between lecturer and students and when the exercise was repeated three months later with a different set of questions, good agreement was obtained (correlation coefficient of 0·80).

Cases 5 and 6 Second Year Physics/Third Year Electrical Engineering

These studies, concerned with a second year physics and a third year electrical engineering course have not yet been completed. A significant feature is that each has raised issues which were not important in the other studies. In one of them, the lecturer was relying on a textbook to provide the backbone of the course, and thus making the lectures into a discussive commentary. The students were finding that they could not follow the chosen textbook easily and the evaluation showed a need for a change of the texts or of the course strategy. In the other course, it became evident that a major problem for the course did not arise from the course itself but from the context, in that students were concentrating on project work which they had to complete and report for assessment over a period when the main bulk of the course was being given.

Conclusions

These studies raise four main issues for discussion. The first, a lesson learnt by those involved as the studies have progressed, is that it is essential to cover a very wide range of topics concerned with each course and its context in order to ensure that important features are not overlooked.

It will also be evident from the examples given that the problems exposed are often specific to the content, aims, teaching techniques and context of any particular course. This second issue has several aspects. The first is that the method of evaluation must be flexible: for example, a standard set of questions for a questionnaire can be a guide but questions may have to be selected, augmented and rewritten so that they are relevant to the particular problems of the course concerned. A second aspect is that simple criteria, eg to label some courses as bad and some as good, are unlikely to emerge from any sensitive evaluation. For example, there is no easy way to compare a course which succeeds because a good lecturer has chosen modest aims with one which causes difficulties because a lecturer – perhaps equally skilled – is trying to stretch the students. A third aspect is that no standard recipe can be supplied for improving courses – each course has specific possibilities and problems. Without more extensive study it is not easy to say how comparisons between one course and another can be helpful in suggesting improvements. The limit of an evaluator's work might be to offer several alternative approaches, he can inform but not override or replace the lecturer's judgement on the matching of his course aims to various teaching approaches and to his own personal style.

A third issue relates to the expertise needed for the studies. It is certain that the questionnaires could be improved technically: care was taken to derive some of the issues explored from the students' views, sometimes using their words and phrases, and to balance the positive and negative questions about a course within its questionnaire as a whole. Many of the questions are leading, but it was found that these were often rejected even in some cases, where they reflected a view agreed in a small discussion group. The other important aspect of methodology was in the role of the evaluator in the group discussions: it was thought very important that he simply facilitate discussion and raise new items for the agenda and not direct or suggest views. For most of the group, discussion was lively and spontaneous and it was not difficult to sustain the required tactics. The use of tape recordings of these discussions was felt to be essential.

A fourth issue is the roles and relationships of those involved. Each evaluation was made by an individual who was outside the department concerned but whose subject expertise enabled him to understand the content of the work he was assessing. At the very least, this enabled his communication with the lecturer and the students to be equally free on all types of issue – those concerned with aspects of the content were not excluded; members of the group concerned were agreed that this was an important feature. A related issue was that each evaluator had no departmental status or role in the eyes of the students and could not threaten a lecturer in the way that an evaluator internal to his department might do. Of importance also was the mutual confidence built up by the close work of the group, helped by the fact that most of those evaluating were also themselves involved as lecturers being evaluated.

These various issues must be considered together in assessing future development.

There is no doubt in the minds of those concerned with the work described here that a great deal was learned about their courses and about ways to improve them which they would not have learned through existing consultation and feedback procedures. It would not in general be either possible or necessary to evaluate any particular course more than once unless serious faults were revealed. However, for one of the courses the evaluation exposed some problems which arose because the lecturer was teaching it for the first time whereas for another the lecturer felt that it provided a stimulus for thinking afresh about a course which he had been teaching for several years.

In considering extension of the work to cover large numbers of courses, the key features are economy and personnel. In respect of economy the experience has been that all of the stages of the procedure adopted were essential. In respect of personnel, the work was not carried out by 'experts' but by interested staff – although it is clear that they could not be described as a 'typical' group of staff. It is possible that with a small amount of training and sample materials, groups of staff could evaluate one another's courses on an informal basis. This informal basis was one of the conditions of the progress of the work and it is a matter for debate whether attempts to promote such work by incorporating it into institutional procedures could extend it without losing the informality and mutual confidence between those involved.

Set by the Pageprint computer filmsetting system and printed offset litho by Page Bros (Norwich) Ltd.